INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.

University Microfilms International

300 N. ZEEB ROAD, ANN ARBOR, MI 48106
18 BEDFORD ROW, LONDON WC1R 4EJ, ENGLAND

7921676

HITCHINGHAM, EILEEN ELIZABETH
SEARCH ASSESSMENT CHARACTERISTICS EXHIBITED BY FACULTY AND STUDENT USERS OF THE MEDLINE INFORMATION RETRIEVAL SYSTEM, WITH AN EXAMINATION OF THE RELATIONSHIP BETWEEN THE SEARCHER/USER INTERVIEW AND THE USER ASSESSMENT OF SEARCH RESULTS.

WAYNE STATE UNIVERSITY, PH.D., 1979

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ✓.

1. Glossy photographs _____
2. Colored illustrations _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Print shows through as there is text on both sides of page _____
6. Indistinct, broken or small print on several pages ✓ throughout _____
7. Tightly bound copy with print lost in spine _____
8. Computer printout pages with indistinct print _____
9. Page(s) _____ lacking when material received, and not available from school or author _____
10. Page(s) _____ seem to be missing in numbering only as text follows _____
11. Poor carbon copy _____
12. Not original copy, several pages with blurred type _____
13. Appendix pages are poor copy _____
14. Original copy with light type _____
15. Curling and wrinkled pages _____
16. Other _____

SEARCH ASSESSMENT CHARACTERISTICS EXHIBITED BY FACULTY AND STUDENT USERS OF THE MEDLINE INFORMATION RETRIEVAL SYSTEM, WITH AN EXAMINATION OF THE RELATIONSHIP BETWEEN THE SEARCHER/USER INTERVIEW AND THE USER ASSESSMENT OF SEARCH RESULTS

by

Eileen E. Hitchingham

A DISSERTATION

Submitted to the Office of Graduate Studies,
Graduate Division of Wayne State University,
Detroit, Michigan
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

1979

MAJOR: EDUCATION (EVALUATION AND RESEARCH)

APPROVED BY:

Maureen A. Sie 5-11-7
Advisor Date

Miriam T. Larson

Donald R. Marquette

Joseph A. Labuta

Acknowledgment

Many people have been of assistance in this study. My most particular thanks are extended to the participating searchers at the Shiffman Medical Library of Wayne State University, the Medical Center Library of the University of Michigan, and the Jesse H. Jones Library of the Houston Academy of Medicine - Texas Medical Center Library. Their suggestions in the preliminary stages were most valuable; their cooperation in incorporating study procedures into their already busy schedules is most appreciated. Special thanks are due to the directors of these libraries for fostering an atmosphere which encourages cooperation in research efforts such as this one. For her immediate interest in the project and her assistance in making long-distance data collection run smoothly, I am particularly grateful to Sara J. Jackson, Assistant Director for Public Services, at the Jesse H. Jones Library.

Mary Kitto of Oakland University has been involved in this study since its inception. I am grateful not only for her technical assistance in setting-up forms, arranging meetings and following through on many details, but also for the very special interest and support she has provided throughout the study.

The work of Susan Woodford and Mary Pensyl of the MIT Science Library in coding the interviews was most valuable.

Their expert knowledge of the data base and the interview situation facilitated the process throughout. Margaret K. Park of the Office of Information Systems, University of Georgia was most helpful in providing copies of transcripts for practice coding sessions.

The participation and contributions of Dr. Joseph Labuta and Dr. Donald Marcotte as members of my dissertation committee are deeply appreciated. Because of her special knowledge of MEDLINE I am particularly grateful for the suggestions and guidance provided by Professor Miriam T. Larson as a member of the committee. My advisor and dissertation director, Dr. Maureen Sie, has been a source of constant encouragement, support, and advice in bringing this study to completion. I am most appreciative of her active and sustaining interest in this effort.

Special thanks are extended to Rita Edwards for her work in typing this report.

For financial assistance in carrying out this study I am grateful for support provided under Grant No. G007702309 from the Office of Education, Office of Libraries and Learning Resources, Department of Health, Education and Welfare.*

---

*The opinions expressed herein do not necessarily reflect the position or policy of the Office of Education, and no official endorsement of the U. S. Office of Education should be inferred.

Finally, my deepest appreciation and gratitude are reserved for my husband Clay. His support has been the constant factor most influential in providing the atmosphere conducive to undertaking this study.

Table of Contents

List of Tables

List of Figures

Chapter 1

Chapter 1 includes a brief introduction, the problem rationale, a statement of the problem, the general hypotheses, the research hypotheses, definitions of terms, and a note concerning the organization of subsequent chapters.

Introduction

Williams (1977) noted phenomenal growth from 1974 to 1977 in the number of on-line searches in data bases accessible through commercial vendors in the United States. During that period the number of searches almost tripled, from 700,000 in 1974 to 2,000,000 in 1977. Among several factors she finds significant in supporting this growth are increased use of the National Library of Medicine (NLM) system; the emergence, in 1977, of a new on-line vendor, Bibliographic Retrieval Services (BRS); increased exposure of students to on-line systems in university and short courses; and decreasing costs per data base record searched.

Interest and activity related to the provision of on-line search services is high. The Reference and Adult Services Division of the American Library Association has formed a special discussion group on machine assisted reference services ("Librarians on MARS," 1974). A conference on The Online Revolution in Libraries, scheduled by the University of Pittsburgh, and for which 300 attendees were anticipated, drew more than 700 interested librarians and library educators (Nyren,

1978). The advent, in 1977, of two new journals devoted to on-line access to bibliographic data bases, ONLINE and On-Line Review, provides support for the belief that the growth of on-line users will be sustained.

Since it is likely that on-line services will become more common in libraries and information centers in the future, it is appropriate to examine the effectiveness of these services in meeting information needs. Within the confines of this study effectiveness is considered to be measures of user assessment of search results in three areas: relevance, concern for recall, and overall value of the search results. Variables to be examined in relationship to the user assessment of effectiveness are interrelationships of user measures, user characteristics, and the user/searcher interview attendant to on-line searches.

Problem Rationale

To date there have been relatively few reports of user assessment of search results for operational, as opposed to experimental, on-line systems. In many of these studies (e.g., Benenfeld, Marcus, Pensyl, & Reintjes, 1975; Hoover, 1976; Kobelski & Trumbore, 1978; Lancaster, 1972; Mick, 1977; Moll, 1972) the user was asked to indicate an overall assessment of search results (satisfaction, usefulness, value), and to respond to more particular aspects of the search (e.g., relevance, recall, improvement over manual searching). Tagliacozzo (1977) suggests that some dissonance may exist between overall user

assessments of searches and the responses to more focused questions concerning search results.

While user status and purpose are reported in many of the studies, there has been no direct examination of differentiation by user status and purpose on the effectiveness measures presented. Rees and Schultz (1967, Vol. 1, p. 179), in an experimental study of relevance (defined groups of individuals reviewing the same test documents for relevance), indicate a basis for believing that student users, as a group, may be less stringent in relevance assessments than scientists engaged in research.

At the present time there are no reported studies of operational on-line systems which have examined the searcher/user interview interaction in relationship to the user assessment of results. It is a common practice for users desiring an on-line search to work with an intermediary (searcher) to obtain the bibliography (Wanger, Cuadra, & Fishburn, 1976). In an interview between the searcher and the user, the subject content and the scope of the bibliography are explored. Since the interview serves as the main vehicle for communicating the user's need to the searcher, searcher/user interaction is considered to be an important related variable in the user assessment of search results. Carmon's study (1975) of the searcher/user interface categorized 51 events which occurred in some, or all of 44 taped interviews for searches of a batch-mode information retrieval system. For many events there

was no distinction made as to the source of interaction (i.e., searcher or user), and there was no direct examination of the relationship between these events and the user judgment of search results.

This study is directed to an examination of three areas for on-line searches: the relationship of specific responses in the user's assessment to an overall user designation of search value; the examination of user characteristics which may be important in interpreting user assessments; and an examination of theoretically significant roles of the searcher and the user in the interview process.

Problem Statement

The problems to be investigated in this study are specified within the framework of four questions.

1. Are user assessments of relevance, satisfaction with the proportion of relevant citations retrieved, concern for recall, and the search interview related to the user assessment of the value of the search in meeting the need prompting the search request?

2. Do specified user groups differ in their assessment of search parameters and in participation in the search interview?

3. In the search interview, is information-giving behavior on the part of the user related to the user assessment of results?

4. Is information-giving behavior on the part of the user enhanced by question-asking behavior by the searcher?

General Hypotheses

In this section three general hypotheses concerning the problem investigated are delineated. These hypotheses concern interrelationships of search assessment parameters, user characteristics in assessing search parameters, and the searcher/user interview and user assessments of search parameters.

Interrelationships. Even though varying question formats are used, a number of evaluation questionnaires used by institutions providing retrieval services appear to elicit evaluative information from the search user in both specific areas (e.g., relevance of citations, absence of known relevant citations, timeliness, feelings about the interview, amount of time saved), and in a more general overall area (e.g., value of the search, satisfaction with the search, helpfulness of the search). Four question types commonly occurring in these questionnaires are: of the citations retrieved how many did the user find to be relevant; did the user feel that the search missed any relevant citations; what did the user feel about the search interview; and did the user find the search results to be valuable (Daniels, 1978; Hitchingham, Note 1).

If overall evaluative assessments are considered to be represented by questions on previous evaluations which ask the user to respond, on a four- or three-point scale, whether the search was, for example, of Major Value, Considerable Value, Minor Value, No Value (Lancaster, 1972; Mick, 1977); or Very Useful, Of Some Use, Of Little Use, Of No Use (Carmon, 1975);

or Very Helpful, Helpful, Moderately Helpful, Not Helpful (Tagliacozzo, 1977); or Very Satisfactory, Generally Satisfactory, Not Satisfactory (Benenfeld, Marcus, Pensyl, & Reintjes, 1975); it is noted that the majority of users (60-91%) respond by checking the first two most favorable categories. This suggests that users of retrieval systems will generally find that the results are valuable or satisfactory, whatever other assessments they make. From an interpretive and diagnostic viewpoint it is desirable to know how more specific user assessments relate to the overall assessment.

User characteristics. Some users of on-line information retrieval systems approach the system with a greater background knowledge of the topic to be investigated in the search. This background knowledge is the result of a continuing involvement with the specific area of investigation, i.e., the search is initiated to supplement already existing knowledge of literature in the subject area. In his review of experimental relevance studies, Saracevic (1970a, p. 137) concludes that greater subject knowledge on the part of the assessor leads to more stringent assessments of relevance. In the same vein it is expected that greater knowledge of the literature of a subject allows more stringent assessments of recall, i.e., the knowledgeable user can note, and is concerned with, a lack of completeness in search results. Knowledgeable users may also place more stringent demands on the searcher in regard to their assessment of the search interview.

Since, as previously noted, users of information systems tend to react favorably to questions concerning the overall value of the search, it is not believed that knowledgeability of the subject provides a distinguishing factor when the overall assessment is considered. In a similar manner, it is not expected that knowledgeability provides differentiation in the satisfaction with the proportion of relevant citations. This satisfaction level is more likely to indicate a tolerance to sifting the wheat from the chaff in the retrieval results, and thus it would appear to be more related to individual temperament than to knowledgeability.

In this study faculty members who use an on-line service for research purposes, where research is considered to relate to grant activity or the publication process, are considered to be literature knowledgeable users. Faculty members can specialize in a limited area, and thus can be more familiar with that area. It is generally assumed that there is some pressure, if only in the tenure process, for faculty members to publish. Such pressure constitutes a job-related need on the part of faculty to be literature conscious. For example, in a study of medical school faculty publication rates (Pearse, Flora, Freeman, & Peeples, 1976), the authors note a peak in publications per person, per year, in the early 40's. They attribute this peak to variations on thesis work followed by a decline until a new research area is selected.

In contrast, most students, because of their role as students, have to focus on a number of subject areas. Subject

areas may be related, but there is less opportunity, because of conflicting information needs, to concentrate in depth on specific topics, and to have a cumulative record of involvement with one area. Thus students, in general, are considered to be less literature knowledgeable.

The interview. Among several themes noted in publications concerning the library reference interview process (and by correlation the interview process for on-line searching), an important role of the intermediary (searcher) as a question asker emerges. For example, Francillon (1959) indicates that ". . . the first question often does not express the real intent of the requester. It is often necessary for the librarian to ask other questions" (p. 193). Taylor (1968) states, "Reference librarians and information specialists have developed, both consciously and unconsciously, rather sophisticated methods of interrogating users" (p. 179). In an instructional module for negotiating the reference query, Jahoda (1975, p. 12) indicates that the reference librarian should use open questions in the initial negotiation stages, and employ closed questions at the final stage of negotiation.

It follows that if question asking is an important role for the searcher, then information giving actions are significant in the user's role in the interaction process. Tessier, Crouch, and Atherton (1977) suggest that during the interaction

for planning a computer search "a user will provide an immense amount of information about his requirements, expectations, and compromises" (p. 386).

It is considered that the amount of information given by the user may impact on search results. Since this information may be either voluntary or searcher elicited, it is desirable to assess the relationship between the searcher's asking of questions and the user's giving of information.

Research Hypotheses

Twelve research hypotheses are proposed in support of the general hypotheses. The first hypothesis relates to interrelationships of user assessment measures, hypotheses two through seven concern user characteristics, hypotheses eight through twelve concern searcher and user activity during the search interview, and their relationship to user assessments of the search.

1. User assessment of search value is not related to user assessment of the percentage of relevant citations retrieved, the satisfaction with the proportion of relevant citations retrieved, the concern for recall, or the search interview.
2. Faculty users will assign lower relevance scores than student users.
3. Faculty users will assign lower recall concern scores than student users.

4. Faculty users will assign lower search interview scores than student users.
5. Faculty and student users will not differ in their assessments of satisfaction with the proportion of relevant citations retrieved.
6. Faculty and student users will not differ in their assessments of search value.
7. Faculty users will exhibit greater information-giving activity during the search interview than student users.
8. Information-giving behavior by the users in the search interview is related to the relevance score for the search.
9. Information-giving behavior by the user is related to the recall concern score.
10. Question-asking activity by the searcher is related to information-giving behavior by the user.
11. Question-asking activity by the searcher is a factor which enhances the relationship between information-giving behavior by the user and the relevance score for the search.
12. Question-asking activity by the searcher is a factor which enhances the relationship between information-giving behavior by the user and the recall concern score.

Definition of Terms

The following section defines terms and concepts used in this study. Definitions are operational and/or literature derived.

Faculty user. Participants indicating on the request form that they are faculty (Instructor, Assistant Professor, Associate Professor, Professor), that the search is for their own use, and that the purpose of the search is grant or publication related.

Information-giving activity (user). In a transcribed interview, the sum of user units which are coded in categories 4, 5, and 6 (Gives suggestion, Gives opinion, Gives orientation) the attempted answers categories, on the interview coding form.

Interaction Process Analysis. Bales (1968) defines interaction process analysis as "an observational method for the study of the social and emotional behavior of individuals in small groups" (p. 465). The method centers on the coding of interaction units (verbal and non-verbal) occurring when small groups are involved in a problem-solving task. Units are coded according to a twelve category system. The first three categories are considered to represent positive, social-emotional areas; the fourth through ninth categories represent neutral, task areas; and the tenth through twelfth categories are designated as negative, social-emotional areas. The task area is further divided into questions (Categories 7, 8, and

9) and attempted answers (Categories 4, 5, and 6). In this study only verbal transactions are scored; hypotheses relate to the neutral, task area.

Interaction unit. Bale's (1950) defines the interaction unit as "the smallest discriminable segment of verbal or non-verbal behavior to which the observer, using the present set of categories after appropriate training, can assign a classification under conditions of continuous scoring" (p. 37). In this study the interaction unit is defined as the smallest discriminable segment of verbal behavior. An example of a conversational segment, divided into units with categories noted in parentheses, and who to whom designations noted in brackets, follows:

> Well [(6) 1-0], suppose we move on a little bit now [(4) 1-0]. It takes a little practice on that of course [(1) 1-0], but I think some of us noticed there was a relaxation that came with it [(5) 1-0]. Did you notice a relaxation as you sat there [(7) 1-0]; how did you feel about it, Paul [(8) 1-2]? (p. 96)

Interview. The face-to-face conversation occurring between a searcher and a user of an information retrieval system. It is "a conversation with a purpose" and the purpose involved is "information getting" (Cannell & Kahn, 1968, p. 526). In general the searcher assumes the role of interviewer and the user the role of interviewee, although roles may reverse in segments of the interview where the user focuses

on getting information from the searcher, and the searcher centers on giving information to the user. The intent of the interview is "to arrive at a clear, narrative, natural language statement of the user's information needs and to gather a number of facts and clues to be used to amplify or refine this statement" (Carmon, 1975, p. 4).

Interview coding form. A tally sheet for noting the frequency of verbal interaction events occurring during the interview. The form is divided into the twelve categories of the Interaction Process Analysis scheme with brief definitions of the categories. It is further subdivided to note event occurrence by type of interview participant (searcher or user).

Question-asking activity (searcher). In a transcribed interview, the sum of searcher units which are coded in categories 7, 8, 9 (Asks for orientation, Asks for opinion, Asks for suggestion), the attempted questions categories, on the interview coding form.

Recall concern score. The average (mean) score obtained by summing user responses to four statements concerning recall on the user response form, and dividing this sum by the number of statements responded to (three or more). The statements assess whether the user (a) believes that the search retrieved most of the relevant citations in the data base, (b) is concerned because there is no way to judge completeness, (c) is concerned because the results omitted relevant citations, of which the user was familiar prior to the search, and (d)

considers that fewer relevant citations than those expected were retrieved. Scale values noted by the user for the second, third and fourth statements are subtracted from eleven to conform to the value of the first statement. Lower scores indicate more concern for recall, higher scores indicate less concern for recall.

Relevance score. The score obtained by dividing the number of citations in the search results which the user indicates are relevant to the search question, by the total number of citations retrieved in the search. For cases in which the user gives an estimate of the percent of relevant citations (0, 1-10%, . . . 91-100%) the number of relevant citations is estimated as that number of citations which corresponds to the higher percent marked by the user on the percent scale (e.g., if the user marks the 81-90% scale, the number of relevant citations is the number of citations which corresponds to 90% of the total number of citations retrieved).

Request form. A form filled out by the user after consenting to participate in the study. The form elicits information concerning the status of the user, the purpose of the search, the educational background of the user, the user's awareness of publications related to the search question, the user's reading habits for journals related to the search, whether the user has published an article or articles on the search request, and the user's future continuity of interest in the topic. In addition, it seeks a written statement of

the problem, and a list of recent citations believed by the user to be relevant to the search request.

Satisfaction with proportion relevant score. The scale number marked by the user, on the user response form, in response to a question concerning the user's satisfaction with the proportion of relevant citations retrieved in the search. Responses can range from one (unsatisfactory) to ten (satisfactory).

Search interview score. The average (mean) score obtained by summing user scale indications noted for five questions concerning the search interview on the user response form, and dividing the sum by the number of statements responded to (four or more). The statements assess, on a disagree (one) to agree (ten) scale, whether the user (a) believes that the searcher was knowledgeable concerning the use of the data base for the question, (b) feels that the searcher understood the request after the interview, (c) believes that the searcher understood the user's purpose in initiating the request, (d) believes that the searcher was thorough in exploring all aspects of the search question, and (e) believes that the searcher suggested terms appropriate to the subject of the request. Lower scores indicate lesser agreement to the statements concerning the interview, higher scores indicate greater agreement to the statements.

Searcher. The individual conducting the interview who will actually do the on-line search of a data base in response to a user's information need.

Student user. Participants indicating on the request form that they are students (undergraduate, graduate, or professional school), and that the search is for their own use.

Transcript. A typed copy, divided into Bales' scoring units, of all verbal transactions noted by the investigator when listening to tape-recorded conversations (the search interview) between the searcher and the user.

User response form. An evaluation form sent to all users appropriate to the study. The form elicits responses from the user in four areas: (a) relevance, (b) concern for recall, (c) value of the results, and (d) the search interview.

Value score. The scale number marked by the user, on the user response form, in response to a question which asks the user to give a value assessment of the search results (the ability of the search to meet the need prompting the request). Responses can range from one (no value) to ten (major value).

Organization of Chapters

The following four chapters represent a developmental study and testing of the hypotheses advanced in the first chapter. Chapter 2 is a review of literature related to this study. The literature provides a theoretical basis for the hypotheses considered and supports the use of Interaction Process Analysis as a methodological tool for assessing transcripts of the search interview. Chapter 3 describes the methods and procedures used in the study. It considers the

sampling procedures, subjects (users and searchers), the development of instruments, data collection, the determination of coding reliability, and the statistical methods used. Chapter 4 includes the presentation and interpretation of data collected. Chapter 5, the final chapter, presents a summary and conclusions drawn from this study.

Chapter 2

Review of the Literature

This chapter presents a review of the literature related to the problem being investigated. In content the chapter parallels the three areas considered under the general hypotheses: interrelationships in search assessment parameters, user characteristics, and the search interview. In structure, assessment parameters and user characteristics are treated together, since the studies intermix these aspects; the interview is considered in a separate section.

Normal reviewing sources were supplemented by Saracevic's reviews (1970a, pp. 111-151; 1975) on the concept of relevance and Hawkins' recent bibliography on on-line searching (1978).

Assessments and Users

Although the relevance study conducted by Rees and Schultz (1967) was not an examination of an operational system, but rather an assessment of experimental relevance judgments (i.e., defined groups of individuals making relevance judgments on test documents) it is included in this review because of some suggestive results concerning differences in relevance assessments by groups of people making a relevance decision. As the authors note, "It was assumed that differences in the extent of subject expertise and experience in research would result in variations in conceptualization of the information

need as revealed by differences in the relevance ratings" (Vol. 1, p. 29).

Subjects were 184 judges divided into five expertise-related groups. The groups were comprised of (a) 40 medical experts in the field of diabetes (i.e., MD's, 14 of whom were researchers, and 26 who were involved in patient care or medical education); (b) 29 medical scientists (individuals with a Ph.D. and working in biomedical research); (c) 25 residents (MD's involved in clinical care of patients); (d) 29 second-year medical students, assumed to have limited expertise and practice in biomedical research, and (e) 61 medical librarians with little, or no direct experience in biomedical research (p. 38). The judges evaluated 16 documents for their relevance to a detailed description of a research project concerning sugar transport in the intestine. The dependent, evaluative measures consisted of four questions directly considering relevance (overall relevance, relevance with respect to the formulation of the research problem, relevance with respect to the experimental technique, and relevance with respect to the interpretation of the findings), and one question concerning the overall usefulness of the document with respect to the research.

Although the total study covered a number of aspects of the relevance judgment (e.g., relevance judgments in regard to different kinds of document representation--citation, abstract, full text; and relevance judgments made in response

to hypothesized stages of the research problem--initial formulation, carrying out the experiment, and analysis of results), only those findings summarized in the conclusion which are related to the current study are considered here.

There were significant differences in the mean ratings assigned by judgment groups to the document set. Scientifically oriented groups (e.g., medical experts and medical scientists) assigned lower mean relevance ratings than did less scientifically oriented groups (e.g., residents and medical librarians) when the overall relevance assessments are considered (p. 271). A more refined breakout, analyzed earlier in the study, indicates that the actual ranking, by group, of mean relevance ratings includes, from high to low, medical librarians, medical experts (non-researchers), residents, medical experts as a total group, medical students, medical experts involved in research, and medical scientists (p. 179). Thus, mean relevance ratings appear to become more stringent as one continues along a scale from clinical to more direct research involvement with the problem.

A second finding of interest to the present study is an indication that, at least on a document to document basis, responses concerning the usefulness of the document were highly correlated to the responses to the question concerning overall relevance. Average ratings, however, differed from responses obtained on the overall relevance question (p. 273). The aspect of this finding that is considered pertinent to the

present study is whether a similar high correlation between the relevance figure determined for a search, and the user's determination of the value of the search in meeting the need prompting the request, will occur.

The AIM-TWX study conducted by Lancaster (1972) appears to be a seminal study in the evaluation of on-line searching because it is one of the first to explicate the idea that one of the best indications of the success of a search is the user's subjective assessment of the search's value to him. In addition, the four-point value scale (major value, considerable value, minor value, no value) developed by Lancaster appears to have influenced the decision of subsequent investigations (e.g., Carmon, 1975; Tagliacozzo, 1977; Mick, 1977; Jahoda, Bayer, & Needham, 1978) to adopt the same scale or to adapt the scale in a similar manner for their own needs.

In the AIM-TWX study 48 users conducted their own searches on the Abridged Index Medicus data base which was available for on-line searching at that time. Users were equally divided between those who had never used the system before and those who had used it at least once. Precision values were determined for each search by dividing the number of unique relevant citations noted by the user in the search results by the total number of unique citations printed (precision scores equate to the relevance score determined in the present study). The average precision figure for 45 searches was 63.1%. As Lancaster notes, this figure is higher than the average

precision (50%) achieved in 299 off-line searches studied in his earlier MEDLARS investigation (1968). In the MEDLARS study precision figures are based on user evaluations of the actual documents (not just the citations); in the AIM-TWX study users were given copies of the actual articles if this was necessary for unequivocal relevance assessments.

After reviewing a number of experimental relevance studies concerning the format of the item which is presented to the user for an assessment (e.g., bibliographic information only--author, title, source, or bibliographic information with abstract, or full text) Saracevic concludes that "Relevance judgments for the same article may be expected to differ from titles to full texts" (1975, p. 340). In the specific relevance study mentioned previously (Rees & Schultz, 1967; Vol. 1, p. 162) the authors note that subjects gave higher relevance ratings when presented with the document title only; relevance ratings decreased as more information was presented.

For the AIM-TWX study (and also the MEDLARS study) this may mean that precision figures noted are, in a comparative sense, more stringent than those reported in other on-line studies. In most of the subsequent reports it is often difficult to determine whether precision figures reflect the user's review of the citations produced, on a review of the actual documents.

Recall values for the AIM-TWX searches were also determined. Conceptually, recall can be considered as an estimate

of how well the search recovers all of the relevant citations available in the data base. In this study, after the user conducted the search a trained search analyst conducted a search on the same topic. The estimated recall performance figure was derived by dividing the number of unique relevant citations found by the user by that same number plus the additional unique relevant reference found by the search analyst who conducted the parallel search. For a group of 36 searches the average recall figure determined in this manner was 57.6%. However, as Lancaster notes in the earlier MEDLARS study, trained analysts may exhibit varying levels of recall performance. For example, for ten NLM searchers conducting four or more searches in that study, recall averages ranging from 54% to 74% were reported (1968, pp. 153-54). The average recall level normally achieved by analysts participating in the AIM-TWX study is not known; however, if they are similar to the NLM searchers considered earlier it seems that user estimated recall for AIM-TWX is a percentage of the best that an analyst could do, which, in itself, is likely to be less than maximum recall.

Approximately one-third of the AIM-TWX users considered in the study were M.D.'s (associate and assistant professors, residents, interns and postdoctoral fellows); ten users were third or fourth year medical students. The remaining users included research associates and assistants, a physical therapist, and an executive from a company manufacturing

medical equipment. More than half of the searches appear to have been initiated for a research related purpose (e.g., ongoing research; writing a research paper, review article or book chapter; writing a book or thesis; preparing a grant application).

Perhaps the most interesting aspect of the AIM-TWX study is a consideration of the value that users attached to the search results. Like users in a number of studies following AIM-TWX, the group as a whole is perceived as attaching a high value to their search output, i.e., for the 38 searches evaluated on the four-point scale, 67% were designated as being of major or considerable value.

Moll (1972) reports on the evaluative responses of 62 users reviewing 100 AIM-TWX or MEDLINE searches done at the University of Virginia from August 1971 through January 1972 (MEDLINE searching, providing greater access to the biomedical literature, replaced AIM-TWX searching of a more limited on-line data base during the period studied). More than 300 searches were conducted during the study period, thus the return rate for questionnaires appears to have been approximately 30%. For 82 of the evaluated searches, users indicated that the citations furnished were either "most useful," or of "considerable interest." For 18 of the evaluated searches the citations were considered to be of "little interest" or "worthless." It is noted that a total of 2,082 citations were produced for the 100 evaluated searches. For 55 of the searches

an estimated 50% or more of the citations sent were considered by users to be helpful in their work. It is interesting to note that 27 of the searches for which users estimated that 25% or fewer of the citations supplied were helpful were still considered to be searches that were "most useful" or of "considerable interest." This suggests that, for some users, a low relevance score does not reflect upon the ultimate judgment of an overall value of the search.

In 1974, almost two years after MEDLINE services were initiated at the Calder Memorial Library of the University of Miami Medical School, users were surveyed concerning their evaluation of the service (McCarthy, Maccabee & Feng, 1974). Questionnaires were sent to 350 locatable previous users. These users had initiated more than 1,200 searches in the time-frame considered. Almost 50% of the questionnaires were returned.

While users were not asked to respond to a direct question concerning search value, the authors assume general user satisfaction from the responses noted for three questions. First, even though 73% of the users indicated that their search had been paid for by institutional funding, 44% of the users indicated that they would use the service with the same frequency if they had to pay with their own funds. Second, 62% of the respondents said that they had received about the number of citations they wanted. Finally, users were asked to supply specific criticisms of the MEDLINE service; for almost 60% of the returned questionnaires no specific criticisms were noted.

The authors suggest that there may have been a difference in satisfaction related to the type of interview conducted (phone or in-person), but other than one comment in this area from a user, they provide no data to support this speculation.

Benenfeld, Marcus, Pensyl and Reintjes (1975, pp. 6-1 to 6-3) report evaluations by users of on-line services at M.I.T. Under the auspices of the Northeast Academic Science Information Center (NASIC), an experimental, pilot operation of a computer-based reference search service was initiated at the M.I.T. libraries in November 1973. The user population included academic users (faculty, graduate and undergraduate students, staff), and industrial users. In December 1974 an evaluation questionnaire was sent to all users of the service from the time of its initiation. One questionnaire was sent to each unique user, some users receiving the questionnaire may have had more than one search request. Almost 50% of the questionnaires were returned.

Through NASIC, M.I.T. users had access to a range of data bases in science and technology, the social sciences, and the humanities. MEDLINE services were available through the National Library of Medicine. Approximately one-third of the more than 300 searches conducted for the period March 1974 through February 1975 were MEDLINE searches (p. G-1). Thus it is assumed that the evaluations covered the entire range of data bases available, including MEDLINE, although this is not specifically noted.

Responding users were, in general, considered to be satisfied with the service. More than 90% of the users responding on a three-point satisfaction scale (very satisfactory, generally satisfactory, not satisfactory) indicated that they were either generally satisfied (41%) or very satisfied (50%). In response to a four-point-scale question concerning citation relevance (high relevance, moderate relevance, marginal relevance, no significant relevance) most users of the service indicated moderate (39%) to high relevance (45%) for the citations they received when these citations were considered in relation to their initial problem.

Almost 35% of those responding did not choose to designate their status (e.g., faculty or student), thus it is not possible to examine user responses in relation to status for this study.

Jestes (Note 2, 1974) reports on evaluations of an on-line search service directed to the CAIN (Bibliography of Agriculture) data base. In 1973 a grant to support an experimental service to provide citations from CAIN was awarded to the Library of the University of California at Davis. The report notes implementation procedures and costs as well as user evaluations collected after three months of service operation.

The population served included researchers in departments and laboratories at the campus, and Agricultural Extension Specialists who attended a demonstration at the university. The service was free to users, but they were given an indication of the search cost when they received their results.

Questionnaires were sent with each bibliography; of a total of 237 questionnaires sent, 87 (37%) were returned.

Responses for three of the questions are considered here. Relevance of citations is reported as a total for all citations received by respondents completing the questionnaire. Of a total of 10,333 citations sent, 36% were considered to be directly pertinent to the research problems of the user, 25% were thought to be interesting but not directly pertinent, and 37% were considered not useful (2% of the citations were not accounted for). Even though the percent of directly relevant citations was not particularly high for this study, users gave some indications that they were satisfied with the results. Almost 90% of the respondents indicated that the search was worth the cost ($11.61 was the average cost per bibliography), and 67% indicated that in the future they would be willing to pay for such a service.

Jestes' report is viewed as supportive of the direction of the current study because of the results concerning relevance and perceived satisfaction, i.e., a group comprised, for the most part, of academic researchers indicated a relatively low overall precision score, yet appeared satisfied with the results, insofar as satisfaction can be equated with a willingness to pay for future services.

Hoover (1976) reports survey results from users of on-line services in another academic environment, the University of Utah libraries. In April-May 1975 all users received an

evaluation questionnaire with their results. Of the 76 questionnaires distributed 26 were returned. The survey period occurred almost 15 months after the initiation of services at Utah.

Most respondents were faculty or graduate students. Sixty-four percent of the respondents indicated that they were first-time users. Multiple data bases were searched for a number of the requests. Users were asked to give an estimate of the percent of relevant citations in their results by checking a five-point scale (e.g., 0-20%, 20-40%, . . . 80-100%). Fifty-four percent of those responding indicated that 60% or more of the citations were relevant. All of the respondents answered yes when queried whether the service was worthwhile. Hoover indicates that faculty groups had the highest success (i.e., more faculty members indicated 60% or greater relevance), and suggests that this occurred because they knew most precisely the topics they wanted to search. However, examination of the spread of responses shows that each user group, other than professional researchers, had low and high relevance searches of an almost equal number.

Faculty and student users of MEDLINE were surveyed in a university that did not have a medical school (Hitchingham, 1976). The study covered a period of four months, the winter semester at Oakland University. Twenty-one faculty and students made use of the service, which was provided free during that period. Thirty-six searches in five subject fields were

initiated. Thirty-one of the searches were for a research related purpose (on-going research, writing a research paper or review article).

Evaluation forms were sent with each of the completed searches. Twenty-seven (75%) of the evaluation forms were returned. All of the students returned forms, while only 13 out of 22 faculty searches were evaluated. Of the 27 searches evaluated 25 were found to be of major or considerable value. Users were asked to mark relevant citations on the printout, a precision value for each search was computed after examining the marked printouts. The average precision value for all searches was 53.9%. Precision values achieved for the searches had no apparent relationship to the amount that the user indicated she/he would pay for such a service in the future (i.e., low and high precision values were noted in conjunction with willingness to pay from $2.50 to $50.00 for future searches). Higher mean precision values are noted for student searches (60%) than for faculty searches (47.3%), and it is suggested that this may be indicative of less critical judgment on the part of students as to what is, or is not relevant to their topic.

Tagliacozzo (1977), in a follow-up study of MEDLINE users who had requested a search from one of seven midwestern centers during the period April-September 1973, achieved one of the highest response rates noted in reviewing on-line evaluations. Almost 90% of the 1,017 questionnaires reaching the addresses

were returned. From an earlier study detailing the characteristics of all the MEDLINE users (Tagliacozzo, 1975, p. 295), these responses can be assumed to have come, for the most part, from in-person users of MEDLINE as opposed to requests received by phone, teletype or mail. It also appears that responses reflect both users who had one search, and users with more than one search (i.e., one user may have returned several evaluations). Almost half of the 810 search initiations were first-time users of MEDLINE.

Two questions directed tc users asked that they consider the "helpfulness" of the MEDLINE search and the "usefulness" of search results. For the helpfulness aspect users were asked to indicate whether the search was not helpful, moderately helpful, helpful, or very helpful. For 895 searches evaluated in this item, 60% were considered to be helpful, or very helpful.

The second question asked the user to mark a point on a line to indicate the usefulness of the search results. The line was anchored at the left by the words "completely useless" and at the right by "very useful." The line was divided into seven equal intervals to obtain scores (-3 to +3). Tagliacozzo suggests that this question required more precise judgment on the user's part than the "helpful" question. For the 826 searches evaluated on this question 69% were scored for points corresponding to +1 through +3. There were fewer responses to this question because some responses were omitted due to a typing error on the questionnaire.

When the responses (N = 821) to the two questions were compared by means of a frequency matrix of helpfulness responses versus line scores, the author notes that there were some ambiguities in the moderately helpful and helpful categories. For example, six responses indicate that the MEDLINE search was helpful, but a -1 to -3 score is noted for these same searches on the usefulness dimension. Response patterns noted for "moderately helpful" show an almost normal distribution when plotted against usefulness scores. Tagliacozzo suggests two interpretations of this occurrence, either the word moderately implied positive connotations to some users, and negative to others; or that users marking "moderately helpful" were "those who did not have--or were unwilling to express--a strong opinion on the MEDLINE service" (p. 246).

The first interpretation appears more likely and presents an example of the importance of providing measurement scales which can be clearly interpreted by the user. In this case it seems that the use of moderately, a "middle" or "medium" type word, on a four-point response set with no real middle (the helpful response alternatives) almost forced a spread of differing interpretations by the user.

In addition to the helpfulness-usefulness comparison, the author contrasts the users' indications of the number of useful references retrieved with values noted on the helpfulness dimension (N = 706). Dichotomizing the helpfulness responses (higher = very helpful, helpful; lower = moderately

helpful, not helpful) and the number of useful references indicated by the user (0-5, and 6 or more), she notes a strong association between the number of useful references retrieved and the judgment of helpfulness. More than 80% of the users with the higher numbers of relevant references also indicate that the search was helpful or very helpful; only 40% of those retrieving five or fewer references gave such indications.

An assessment aspect that might be considered to relate to a concern for recall by the user was also studied. For this contrast (N = 795) frequencies of dichotomized helpfulness values were related to the user responses indicating either that no relevant citations were missed, or that one or more relevant citations were known to be missing. She notes a significant association in these measures, i.e., 70% of the responses indicating no missed citations were aligned with upper values on the helpfulness scale; only 49% of the users noting one or more missed citations chose the higher categories.

This study presents some difficulty in interpretation because of the varying N's used in presenting results, and the lack of a clear indication of independence of responses for the chi-square determinations (i.e., in reviewing the number of users and the number of searches it seems that some of the responses result from one user assessing several searches, while others are based on one-user, one search). Nevertheless, it is noted as particularly important because the author clearly suggests that users may give contradictory and dissonant

responses when considering several aspects of an evaluative questionnaire. Furthermore, she suggests caution in interpreting from single responses, particularly those which elicit general or overall judgments, that the user's information needs were, or were not, satisfied by the service.

Relatively few evaluation studies have been reported for users in a non-academic environment. Fosdick (1977) reports the responses of on-line users at the U. S. Army Construction Engineering Laboratory (CERL) in Champaign, Illinois. Evaluation forms were sent to 27 users noted as recipients of on-line searches over a two-year period; 26 of the forms were returned. Fosdick used the same scale for relevance as that previously noted in Hoover's study (1976). Unlike the academic users, respondents at CERL indicated generally lower estimates of citation relevance. Almost 70% of the users indicated that 40% or fewer of the citations were relevant. However, in response to a broader question concerning the usefulness of the service, CERL users evidenced the same favorable trend apparent in many other on-line evaluations. Twenty-five of the respondents indicated that they felt that the service was a useful addition to other library services; twenty users believed that the search was worth the cost.

A second study focusing on non-academic on-line searching is the evaluative component of the final report on the Lockheed-Public Library Experiment, funded by the National Science Foundation (Mick, 1977). The project involved development and

implementation of a program to assess the viability of providing access to major on-line bibliographic data bases in the public library setting. Only those evaluative aspects related to the current study are considered here.

During the first year of the project services were provided free to users of four California public libraries. In the second year one-half of the actual costs (terminal connect time and printing costs) were charged back to the user. Two libraries continued with the project in the third year. During this period the fee system was based upon total recovery of connect and printing costs.

In the first year there were 1,236 patrons who requested a search. For the second year (recovery of half the costs) usage dropped; in this period there were 611 users. In the third year (full cost recovery, only two libraries participating) 326 users requested an on-line search. Students (graduate and undergraduate), educators (teachers, professors, school administrators) and scientists/research personnel accounted for the bulk of users over the three-year project. Use by graduate students and educators increased over the time span encompassed by the project, while use by undergraduate students and scientists declined as half, and full cost fees were introduced.

Although the details of eliciting evaluative information from users are not clearly explained, it appears that a questionnaire was sent to users at some time during the study.

Response rates appear to have been fairly low, i.e., evaluative responses noted represent 36% of the first year users, 26% of the second year users, and 29% of the third year users. Responses to questions concerning the search value, the adequacy of results, and the number of citations considered useful are discussed here.

In general, responding users showed little change over the three-year period in their perception of the value of the search. When asked to make a value designation on a four-point scale (major value, considerable value, minor value, or no value), 68% to 76% of the users over the three years indicated that the search was of major or considerable value.

In response to a yes/no choice concerning whether the results adequately answered the question, 53% to 59% of the users responded in a positive mode. However, Mick does note that if responses for the third year are examined by individual library, there appears to be a sharp division in response modes by users of the two libraries which account for 94% of the evaluations returned in the third year. More than 60% of one library's users felt that the question was adequately answered, only 31% of the second library's users gave such an indication. Mick suggests that the difference in client feedback may be explained by the fact that searchers in the library with the more satisfied users spent a greater amount of time in pre-search preparation, than did searchers in the second library.

Although relevance determinations for citations were not given, in reviewing the results of Mick's study it is possible to speculate that many users were satisfied with a fairly low proportion of useful articles (e.g., almost 75% of the third year users found 0 to 15 citations to be useful). An earlier, interim report of the public library experiment (Summit & Firschein, 1976, pp. 3-3) indicates that in the second year of the study an average of 88 citations per search were printed off-line. Mick reports that 60% of the second year users had 0 to 15 useful citations.

The evaluation of the public library experiment appears to be important because it further substantiates the belief that whatever the other factors involved, users will give rather favorable responses to an overall question concerning the value of the search.

One of the few situations in which on-line users in two different types of organizations were studied has been reported by Jahoda, Bayer, and Needham (1978). The results presented were obtained from records maintained for each search, and user feedback obtained for most of the searches. Participants in the study included 50 individuals from the Chemistry department at Florida State University (faculty, graduate students at the dissertation stage, and research associates), and 234 scientists and technologists at the Monsanto Textile Company. Services were free to users during the 13 month period reported (the study is still in progress,

currently users are assessed half of the actual search costs). Results from 353 FSU searches and 345 Monsanto searches are reported for the free period. Findings related to the present study are examined here.

Results from the two groups are contrasted by chi square determinations. Academic users were found to be more interested in an exhaustive approach to the literature (everything available) than were the industrial users. Monsanto users were more likely to approach the service because they were looking for specific facts and procedures. When the user analysis of results is considered, the authors note that both FSU and Monsanto users were generally pleased. Users were asked whether the number of citations they received was "about right," "too many" or "not enough." For 68% of the FSU searches and 64% of the Monsanto searches, users indicated that they had received about the right number. The average or median number of total citations retrieved for the searches is not reported, although it is known that for the FSU searches only 25% resulted in more than 21 on-line citations being printed, and there were off-line prints for 36% of the searches. Sixty-three percent of the Monsanto searches had off-line citations. It seems that even though FSU users are characterized as being more exhaustive in their approach to the literature, they may have received search results which, on the whole, included fewer citations than those received by Monsanto users. There was little concern by either group that too many citations had

been retrieved (7% of the searches at FSU, 5% at Monsanto); some concern that not enough citations were presented was more likely to be expressed (13% at FSU, 19% at Monsanto). At each institution, approximately 10% to 12% of the searches were not evaluated on any of the user response measures.

There were no significant differences in the two groups when they reported their perception of the overall utility of the searches. The theme of generally favorable responses to an overall evaluative question noted previously in this review, is repeated for both user groups. Users were asked to indicate whether the searches were "very useful," "of some use," "of marginal use," or "of no use." FSU users found that 78% of the searches were either very useful or of some use, at Monsanto 75% of the searches were in these categories.

Student users of on-line services were the focus of interest in a study at the University of Delaware (Kobelski & Trumbore, 1978). The groups studied were satisfied and non-satisfied student users, who had received a cost-subsidized search during the 1976-77 academic year. The period covered was the second year of operation for the search program. The subsidization system meant that students were charged half the normal charges for on-line searching (actual connect costs plus 15% surcharge for paper, phone, etc.). The average cost to subsidized users was $10.31.

One hundred and fifty-four subsidized searches were performed during the year; evaluation forms were returned for

107 searches (69%). As is true for a number of other studies the independence of responses is not reported, i.e., it is unclear whether each search evaluated represents a unique user, or whether some users had multiple searches and thus returned more than one evaluation. Searches were run on a range of data bases, with heaviest use occurring for Psychological Abstracts, Biological Abstracts and ERIC. An average of 1.4 data bases were searched for each request.

The details of the evaluation form are not reported but it appears that a dichotomous choice, satisfied/not-satisfied, was one of the items, since other search parameters are related to this aspect in the report. More than 80% of the searches evaluated were considered to be satisfactory. For 70% of the evaluation forms it was noted that the responding user indicated that she/he would do another computer search in the future.

The authors indicate that the evaluation forms were examined for factors that would account for student satisfaction or dissatisfaction. They conclude that neither the student's status or purpose in requesting the search were related to student satisfaction. They also indicate no relationship between the cost of the search and expressed satisfaction. It is further suggested that no relationship exists between the percentage of relevant citations retrieved and satisfaction, because some searches in which all citations were relevant did not satisfy the user, while other users were

satisfied with less than one relevant citation in ten in their results. Using chi square they indicate a relationship between the number of citations generated and student satisfaction. All students receiving more than 100 off-line citations were satisfied. All students receiving at least 40 relevant citations were satisfied.

The application of chi square to the results reported appears to be questionable, both because of the previously mentioned suggestion that responses may not have been independent, and also because any determinations made appear to relate to after-the-fact data sifting. However, this study is interesting because of the wide range of values reported on the measures (number of citations generated, number of relevant citations generated, percentage of relevant citations generated), which are all aligned with an overwhelmingly favorable response concerning satisfaction.

The Search Interview

Quantitative studies related to the search interview are far less common than those which consider user evaluations of on-line searching. Carmon (1975, p. 6) reviewed the published literature for reference services and computer-based retrieval services, and concluded that the what of library reference work (and by equivalence data base searching) was somewhat defined, but that the how aspect was largely undefined. Taylor (1968, p. 180) called the negotiation aspect of a reference interview a complex act of human communication, with

the user trying to describe for another individual not what he knows, but something he does not know. At that time he suggested that quantitative information about the process was non-existent. More recently, Lynch (1977, p. 11) has confirmed that there has been little empirical analysis of the reference interview. Carmon's study (1975) is most related to the current investigation. Other reports, peripherally related, are also discussed in this section.

In the MEDLARS study mentioned previously in this review, Lancaster suggested that 25% of the recall failures, and 16.6% of the precision (relevance) failures could be attributed to defective user-system interaction (1968, p. 99). He mentions two basic types of interaction failure:

> 1. The situation, long known to librarians, of the user who puts an imperfect request (i.e., a request that does not precisely match his information requirement), or the situation in which the information need is captured imperfectly by a librarian or search analyst.
>
> 2. The situation in which the user puts a request that is a fair reflection of his information need, but recall failures result from the fact that he is not fully aware of the types of article that exist and could be of use to him. (p. 101)

He notes that failures of the second type account for 20% of the inadequate interaction searches, and that this problem could only be solved by a browsing or iterative search (i.e., the user views some citations, and thus expands, narrows, or refocuses his request).

One of the more interesting findings related to interaction in the MEDLARS study was an examination of precision

and recall figures by mode of interaction (personal, the user talks to the search analyst; local, the user talks to a librarian who may or may not interact and who then transmits the request to a search center; and no interaction--mailed requests). Personal interaction appeared to lead to slightly depressed performance on both measures. Lancaster suggests that this may have occurred because analysts discussed the search request in terms of the vocabulary available for searching, rather than in the context of the user's actual information need. One of the recommendations resulting from the MEDLARS study was that a new user request form be designed which would more fully explicate the framework of the user's information need (e.g., the user would be asked to submit a narrative statement of need, indicate the purpose of the search, list known relevant documents, indicate any limiting factors that might apply to the search, and indicate tolerance levels in regard to relevance and recall). This recommendation appears to have had a far-reaching effect. Not only were MEDLARS forms subsequently revised, but it appears that Lancaster's suggestions were adopted as a prototype for a number of request forms developed for on-line searching (e.g., see the composite user request form developed by Daniels, 1978).

A somewhat similar occurrence (slight depression of performance factors related to user-analyst interaction) is briefly discussed by Saracevic (1970b, pp. 677-678). The

report summarizes an experimental project carried out at the Center for Documentation and Communication Research at Case Western Reserve University. Part of the study concerned question analysis and searching of real queries submitted by users in the field of tropical medicine. Requests were processed against an experimental data base of 600 documents. Strategies were developed, and results were examined for each of five different search mechanisms, (a) using terms from the submitted question only, (b) terms from the question plus terms from the thesaurus, (c) terms from the question plus terms from any source other than the thesaurus, (d) terms from the question plus terms from the thesaurus plus terms from all other sources, and (e) consultation and verification by the user of the most complete strategy (terms from question, thesaurus, and other sources). A slight drop in the number of relevant citations retrieved occurred at this last stage. The author attributes this to the deletion by the searcher of some of the expanded terms after consultation with the user. He suggests that this does not mean that contact with the user is not beneficial, but rather that the way in which it was conducted was not successful.

Link notes contrasting results when the performance measure utilized is some aspect of user satisfaction with a computer-based information retrieval system (1972, p. 51). Two user groups were exposed to two different methods for obtaining searches from the ERIC data base. The experimental

group formulated their own search strategies by utilizing a CAI training package. The control group obtained their searches through the normal mechanism, i.e., interaction with a search intermediary. On four satisfaction measures (number of hits, relevancy of hits, amount of time spent, general satisfaction), the control group (those working with an intermediary) indicated slightly higher scores than did the experimental group. The author notes, however, that t-test analysis showed the differences to be non-significant.

Considering the Lancaster and Saracevic findings we are confronted with a situation that suggests that searcher-user interaction may sometimes operate at a less than optimal level in regard to certain evaluative measures applied to searches. At the same time, most operational services rely on the interview mechanism for question translation (user to searcher to system), since it appears that end users are not particularly interested in doing their own searches (Lawrence, Weil & Graham, 1974, p. 368; Wanger, Cuadra & Fishburn, 1976, p. 194). A partial solution suggested by Lancaster is to elicit more information from the user by means of an expanded request form, Saracevic believes that the interview may be beneficial if it is conducted successfully. Some implied or prescriptive attributes of the successful interview are considered in the following literature.

Taylor's five filter model (1968) of the reference interview was first proposed in the late 60's. He conducted a

number of interviews with special librarians and information specialists to determine their methods of question negotiation. He states that this group was selected because they are usually concerned with substantive questions, receive inquiries from motivated and critical users with pre-knowledge of what is acceptable as an answer, and are familiar with negotiation techniques. Based upon information obtained in these interviews he suggests five steps in negotiation, from which the librarian selects data useful in conducting the search. The five steps include:

1. determination of [the] subject; [of the request]
2. objectives and motivation; [of the user]
3. personal characteristics of [the] inquirer;
4. relationship of inquiry description to file organization;
5. anticipated or acceptable answers. (p. 183)

Each step implies that the librarian will assume a question-asking role and that the user's role will focus on information giving.

Videotape scripts illustrating "poor" techniques of query-negotiation and elements of a "well-negotiated" query are included in Jahoda's instructional module for answering reference questions (1977, Appendix D). The situation is supposed to simulate reference desk encounters in an academic library. The poorly negotiated information request involves nine questions asked by the librarian. In the well-negotiated request the librarian asks 15 questions.

Somerville (1978) takes a more prescriptive approach to the search interview by including a list of 20 things the

search intermediary should do during the interview. She suggests that the searcher:

> Ask questions of the user to ensure your understanding of the subject . . .
> Determine if the user prefers a comprehensive search or a narrow one . . .
> Make sure that you identify all the restrictions that the user wants placed on the search strategy. (p. 23)

Some correspondence with the five filter model is evident.

Implications and prescriptions for query negotiation have gone beyond the purely mechanistic aspects of the interview. They often include not only what should occur (the transfer of information) but also the manner in which it should occur. For example, Jahoda's list of good and poor ways to negotiate a request suggests that in the good negotiation the librarian will make eye contact with the user, give the user full attention, make the user feel at ease, show empathy for the user, and be aware of non-verbal cues (1977, p. 25). Somerville's list of "DO's" notes that the interviewer should make the user comfortable in discussing information needs by utilizing interviewing and counseling techniques (p. 23).

Gothberg (1975, 1976) addressed the non-mechanistic aspect of the reference interview in her study of the effect of immediate and nonimmediate verbal-nonverbal communication behavior by librarians on user satisfaction. Two reference librarians were trained to display immediate or nonimmediate verbal-nonverbal communication during a reference interview

in a public library. Immediacy indicates a quality of liking or closeness in an interpersonal relationship. Each librarian was involved in immediate and nonimmediate reference interviews. Reference transactions were recorded and videotaped. Library users (60) were selected on the basis of their availability in the library, need to negotiate the reference question and willingness to answer the questionnaires.

After the interview the investigator approached the user and asked her/him to indicate, by means of a questionnaire, satisfaction with the reference interview, satisfaction with the user's own performance in negotiating the reference question, and satisfaction with the transfer of information. Verbal and non-verbal components of the recorded interviews were coded for immediacy and nonimmediacy. An $ANOVA_2$ computer program was used to analyze data.

User satisfaction with the reference interview and their own performance in participating in the reference interview were significantly related to interviews with immediate communication. However, when the user's satisfaction with the <u>transfer</u> of information was considered there was no significant relationship observed between immediacy in interviews and this satisfaction measure. The author observes that lack of effect in this area may have meant that displays of verbal and nonverbal liking were not sufficient to bring about the trust necessary for users to divulge their true needs and lack of knowledge about the library and its tools.

Gothberg's study is interesting for several reasons. First, it appears that some defined aspects of user satisfaction may relate to the sociobehavioral qualities of the reference interview, while another measure (satisfaction with the transfer of information) may not. Second, the study is of methodological interest because it is one of two studies isolated which incorporates observations of a real interview situation, quantifies the observations, and assesses the relationship that exists between the quantified observations and a user satisfaction measure. Finally, in its focus on the social-emotional aspects of the interview, it provides a direct contrast to the present study which focuses on a task area of the interview (information-giving and question-asking).

Lynch's examination (1977, 1978) of reference interviews in public libraries is more like the current study of the search interview for it also focused on the mechanistic aspects of the interview. However, user satisfaction was not considered in the study.

Lynch posed eight questions:

1. How often does a reference librarian interview the patron who presents a reference query?

2. Does this frequency vary according to the type of transaction involved?

3. Are interviews more frequent when the librarian is less busy?

4. When an interview does occur, what gross categories or levels of information are sought by the librarian?

5. How often are the questions of the librarian open questions and how often are they closed questions?

6. Does the reference librarian use the secondary questions (probes) used by other interviewers?

7. How does a librarian discover that the query a patron first presented is not the query he/she wants answered?

8. How many primary questions does the librarian ask the patron in an interview? (1977, p. 43)

Reference interviews in four public libraries were recorded. Of the 366 interviews recorded, 309 were ultimately transcribed. Tapes were not transcribed in their entirety. Since question asking by the librarian was the focus of the study, only questions that the librarian asked were transcribed, along with enough additional material to make them meaningful. Transactions were classified according to their nature, i.e., directional, holdings transactions, substantive or moving. Interviews were considered to occur in those cases in which the librarian asked the patron one or more questions. The investigator developed an 11 category scheme for analyzing holdings' transactions, and a 20 category scheme for substantive transactions. In addition, questions were analyzed on an open/closed scale, and a scale to assess probing questions by the interviewer. Some problems in reliability of coding were noted. Findings related to questions five, six and eight posed by the investigator are considered here.

Open questions, i.e., those which allow flexibility in user response, were employed infrequently in the interviews.

In holdings and substantive interviews they comprise eight percent of the questions asked. Ninety percent of the questions were closed. Two percent were considered to fall into an intermediate category.

The author notes that the sixth question concerning secondary or probing questions by the librarian could not be resolved (1978, p. 32). In coding there was difficulty in establishing whether a true probe occurred (a librarian tries to elicit more fully information about an aspect of the patron's information need), or whether the question by the librarian was actually an attempt to be sure that the patron's original question had been heard correctly.

Primary questions (questions by which the librarian introduces some aspect of the patron's search for information and asks for content new to the interview) were infrequently used in the interviews. Fifty-two percent of the interviews involved only one question. Eighty-nine percent of the interviews had three or fewer questions.

Conceptually this study is quite interesting in its attempt to make direct observations of actual reference interviews; however, the results are extremely difficult to interpret. Interpretation problems appear to be associated with the multiplicity of categories and classification schemes applied to the questions, and concomitant problems arising in the area of reliable classification.

The final study reviewed here (Carmon, 1975) has had a direct influence on some of the aspects of the present

investigation. Carmon's study had two goals; first, to collect descriptive and quantitative data concerning the reference process for computer based literature searching, and second to develop a model of the user interface for a projected network model of system use.

The study was conducted at two sites, the University of Georgia and UCLA. Each site provided batch-mode searching of a number of data bases. Georgia's system allowed for both current awareness searching (SDI) and also retrospective searching of earlier years on the data bases. UCLA provided current awareness searching only. Evaluative components in the Georgia results do not make a distinction as to whether retrospective or current awareness searches are being considered.

At the University of Georgia interviews were conducted by four reference specialists; these specialists were by job function involved almost exclusively with computerized literature searching. UCLA results originate from searches profiled at a number of institutions in the California system, with the actual computer run being accomplished at UCLA. Searchers in the California system were more likely to have several other job responsibilities in addition to computerized literature searching.

The University of Georgia results cover a period of five months during which 333 users initiated a search request. Approximately two-thirds of the users were graduate students,

although some overlap in status categories is noted. At UCLA graduate students comprised almost half the population studied. Twenty-five percent of the interviews at Georgia were recorded, and ultimately 44 of these interviews were transcribed. During the same period UCLA processed almost 150 search requests. Fifty-nine of the UCLA interviews were recorded and transcribed.

User evaluation forms were sent to University of Georgia and UCLA users. Twenty-five percent of the total University of Georgia users returned an evaluation form. For searches with recorded interviews there was a return rate of 77%. At UCLA 52 user evaluation forms were returned.

On a general satisfaction measure (very useful, of some use, of little use, of no use) responding users at both sites indicated that they were satisfied with the results. At Georgia 96% of the searches that were taped were found to be very useful, or of some use; for all the returned evaluation forms 89% of the searches clustered in these categories. At UCLA 91% of the returned evaluation forms indicated that the user found the search to be either of some use or very useful. At both sites the most frequently ranked specific dislike expressed about the service concerned recall, i.e., users ranked no way to judge completeness high on the disliked features list.

Transcribed interviews were coded at each institution. Slightly different, self-developed coding systems were used

at each site. Frequencies for events were noted by the occurrence or non-occurrence of specific events; these events were not ultimately related to the users expressed satisfaction measures. For both sites it was determined that the events coded were likely to occur throughout the interview. For example, the searcher might focus first on question negotiation, inject some descriptive information about the system, do some strategy development, and return to more question negotiation.

The low frequencies noted for a number of categories in the transcribed interviews were of interest. For example, at Georgia there were no instances in which the searcher asked the user to restate the question, an occurrence which might elicit additional information. Asking the user whether any other concepts should be added to the search occurred in only 10 of the 44 transcribed University of Georgia interviews. The users were asked to give an estimate of the available published literature in nine interviews; relevant author terms were sought from the user in thirteen of the interviews. Language requirements of the user were directly sought in ten interviews; the user was asked to confirm the strategy in only three of the interviews. If, as suggested in previous writings on the interview process, question-asking is an important role for an interviewer, it seemed that an examination of the magnitude of question-asking as related to user information-giving

and the ultimate assessment of results would provide an interesting framework for teaching the interview process, and also practical application in real search interview situations.

Because of coding difficulties noted in several of the previously reviewed studies it seemed important to separate the problem of identifying a classification system from the actual treatment of the items being examined. In other words, it did not seem desirable to invent a unique classification scheme for question-asking and information-giving but rather to utilize an existing scheme with capabilities for replication. To this end the literature of content analysis was briefly reviewed (Berelson, 1954; Borgatta & Crowther, 1965; Holsti, 1968, 1969). The system selected for this study will be more fully treated in the following chapter on methods and procedures.

Chapter 3

Methods and Procedures

This chapter considers methodological and procedural aspects of the study which establish the perspective for reviewing results. Specifically discussed are (a) the research design, (b) limitation of the study to a particular data base, (c) the environmental constraints which constitute the framework from which subjects were selected, (d) data collection instruments employed in the study, (e) data collection procedures, and (f) data treatment methods utilized in generating measures appropriate for describing the results, and examining the operational hypotheses.

Design

The study is characterized as descriptive research focusing on three areas. These include the examination of interrelationships of user assessment areas often included in evaluating schemes applied to operational on-line search services; the examination of differences in user response styles exhibited by defined subsets of the universe of users who utilize on-line search services; and the examination of relationships between specified events occurring in the search interview and the user assessment of search results.

Data was collected in a natural setting, that is, collection was incorporated into the normal pattern of procedures

employed in responding to a request by a user for an on-line search. Artificial aspects introduced in this study included asking the user to complete a request form, tape-recording of the search interview, and requesting that certain users participate in an evaluation of search results (questionnaires mailed to users).

Interviews were transcribed, and quantified by means of a particular content analysis scheme (Bales, 1950) applicable to the questions examined in this study.

On-line Data Base

This study was limited to interviews and searches conducted for a particular on-line data base, MEDLINE. Several factors supported this choice. First, the MEDLINE data base has been widely available for searching for a relatively long time (McCarn & Leiter, 1973) in contrast to some of the other on-line data bases available through the commercial vendors. Because of this, it offered the opportunity for investigation in an atmosphere uncomplicated by variables associated with the use of more recently available data bases (e.g., few users available for study because they are as yet unaware of a new service; searchers interviewing for, and searching on a data base which is unfamiliar to them). Second, because of its broad subject content MEDLINE is often applicable to requests in a number of disciplinary areas (medicine, biology, psychology, some social sciences); interviews for MEDLINE

may mirror to some extent the interviews employed in searching other data bases. Third, certain aspects of question negotiation for MEDLINE searching (thesaurus use, strategy development, establishment of preferences for time coverage and citation format, etc.) were likely to provide a common denominator in interviews at all of the search sites; furthermore, these aspects are comparable to those utilized in question negotiation for several other data bases. Finally, an earlier experimental study of relevance judgments focused on assessments made by groups of individuals in the biomedical field (Rees & Schultz, 1968) and established, within this context, a basis for believing that depth of research knowledgeability may be associated with more stringent relevance assessments. An examination of this aspect in an operational situation was of interest in the current study.

Thus, the limitation to an examination of MEDLINE searches is not to be viewed as a specific attempt to evaluate the MEDLINE system, but rather MEDLINE was chosen for study because it is a data base which has many features common to other data bases, and MEDLINE users were likely to be similar to users examined in an earlier study.

Sampling Plan

Subjects were in-person users and searchers at three institutions providing MEDLINE searches. The following sections detail characteristics of sites, searchers, and users which define the sampling framework.

Sites

Collection sites included the Vera Shiffman Medical Library at Wayne State University in Detroit; the Medical Center Library at the University of Michigan in Ann Arbor; and the Houston Academy of Medicine-Texas Medical Center Library. The first two libraries were included as participants because they had expressed interest in cooperating in the study, and their locale was convenient for regular visits by the investigator to pick up tapes and other materials. The third library was included when data collection appeared to be progressing at a rate slower than anticipated. Materials from Houston were mailed to the investigator. All institutions, searchers and users were promised anonymity in the reporting of final results.

The Shiffman Medical Library serves both the immediate population of faculty, students and researchers associated with Wayne State University, and a more diverse group of urban and hospital-related users with information needs related to the health care field. Shiffman is also the Regional Resource Library for Kentucky, Ohio and Michigan (KOM).

The Medical Center Library at the University of Michigan is located in the hospital complex of the University Medical Center. The Library serves the information needs of researchers, faculty, students, and clinicians affiliated with the Medical School and other health profession schools of the University. It also serves as a resource for

individuals associated with a number of prominent research institutes at the University.

The Houston Academy of Medicine-Texas Medical Center Library is a consortium library governed by representatives from major participating institutions--the Houston Academy of Medicine, Baylor College of Medicine, Texas Women's University, the University of Texas at Houston, and Texas Medical Center, Inc. It also serves as the Regional Resource Library for the South Central Regional Medical Library Program.

All of the libraries charge for MEDLINE searches, with fees ranging from $5.00 to $15.00 or more, depending on the number of backfiles searched and the format of the printouts (e.g., inclusion of abstracts). One library prefers to have the user present for the interview and terminal search, another library intermixes these procedures (some interviews are combined with the search, some searches are run after the user leaves), and the third library, in most cases, separates the two procedures (the search is run sometime after the interview with the user). However, some terminal searches with the user present were noted as occurring in all of the libraries for this particular study.

Searchers

Eleven searchers participated in the study. All searchers can be considered to be relatively familiar with MEDLINE since they had conducted 125 or more searches in the 12-month period prior to the study. One searcher had at least a year's

experience with MEDLINE; all others had two or more years of experience. Undergraduate majors of searchers were predominantly in areas other than the physical or life sciences. Two searchers had undergraduate degrees in the sciences.

Searchers indicated that training for MEDLINE searching had been accomplished in a variety of ways. Four searchers had attended training sessions of varying lengths at the National Library of Medicine; the others were trained for MEDLINE searching through several mechanisms (singly or in combination). These included NLM sessions scheduled in their vicinity, use of the MEDLEARN training sequence, training by a more experienced colleague, and self-teaching.

Users

Several limiting factors, either external to the study or design imposed, determined the sample framework of the potential pool of users investigated in this study.

The pool from which users could be drawn included all unique, in-person users requesting a MEDLINE search at one of the search sites during the data collection period. Thus, excluded from the outset were search requests received by mail or phone, search requests from a user who had already participated in an earlier session, and search requests from users who asked for another data base even though MEDLINE may have been subsequently used to satisfy the information need (all of the sites provided a range of search services on data bases other than MEDLINE).

A second limiting factor on the potential pool of users was self-selection of individuals for participation in the study. In recognition of concerns for user and searcher privacy in regard to tape-recording of interviews, and in compliance with guidelines from several committees charged with monitoring all studies which involve human experimentation, a consent form for participants was developed (Appendix A). Searchers were asked to note the reason for not recording for all searches they conducted during the data collection period. Known decisions not to participate (either searcher or user) occurred in a ratio of less than 1:2 when considered against decisions to participate. The ratio may be higher since less than complete information was available from one data site. In several instances searchers noted that the user did not wish to participate because the user was "in a hurry" and did not wish to take the time to complete the request form. In almost half the non-participating occurrences the decision not to record was searcher initiated and appeared, in a number of instances, to be related to lack of immediate access to the tape recorder or general lack of time to complete all the procedures. Most instances of lack of time for procedures occurred at one site which had a relatively high volume of searches during the data collection period.

A third, design-imposed factor was the ultimate selection of participants to receive evaluation forms. Subjects receiving evaluation forms were selected from the total group of

of recorded participants according to design criteria, i.e., subjects selected had indicated that their primary status was either faculty (instructor, assistant professor, associate professor, professor) or student (undergraduate, professional school student, or graduate student). Furthermore, subjects indicating either of the previous status conditions had to have also indicated that the search was primarily for their own use. This was to preclude a situation in which the ultimate evaluator of the search was someone other than the person taking part in the interview.

Faculty subjects selected to receive evaluation forms were further restricted in that their stated purpose in obtaining the search had to be research related (grant project, preparing an article or review for publication, on-going research leading to publication). This restriction was imposed to parallel the conditions noted by Rees and Schultz (1967) in their experimental study of relevance in which it was noted that knowledgeability and involvement in research were related to more stringent relevance assessments. In other words, it was not desired that faculty searches of a more sporadic or superficial nature (e.g., search for immediate clinical application, background material for a speech, material for class instruction) be included in the evaluation.

A final situational restriction was imposed after-the-fact. In a few cases two users participated in the interview. Since the search interview is considered to be primarily a dyadic

communication process, these interviews would be misleading if analyzed for information-giving behavior by the user. Furthermore, they would present problems in determining the individual most appropriate for evaluating the searches. Searches of this type were excluded.

Data Collection Instruments

Data collecting instruments for this study included a user request form for collecting background information concerning the user, tape recordings of interviews, user evaluation forms sent to participants meeting study criteria, and coded transcripts of selected interviews.

User Request Forms

The user request form (Appendix A) was developed for two purposes. First, to obtain primary information about each user which allowed a decision to include or exclude a particular user from the evaluation process. Questions 7, 10, and 11 (ultimate user of the search, status of requester, and purpose of the search) fulfill this function. Second, for comparative purposes it was also desirable to more fully describe background characteristics of the user that might clarify the framework from within which the user made an evaluative decision. Questions 6, 12, 13 and 15 (number of previous on-line interviews, degrees completed, familiarity with and sustaining interest in the search topic, and ability to list relevant citations) were of this type. Question 14 (narrative search

statement) was included for comparison with an earlier study (Carmon, 1975) in which it was noted that users supply fairly brief statements concerning their search topic. Other questions on the form were designed to accommodate information usually included by one or all of the search sites in their already existing forms.

Tape Recordings

Tape recordings were utilized to capture all verbal interaction between searchers and users. They were turned on as soon as the consent form was signed and turned off at the completion of the interview.

User Evaluation Form

A standardized, generally accepted instrument for electing user evaluations of searches does not exist. The form used in this study is self-developed, but it reflects closely areas considered in other evaluation schemes (Daniels, 1978; Hitchingham Note 1).

Specifically, the evaluation form (Appendix A) was designed to elicit information from the user in five areas: (a) the relevance score for the search, where relevance is determined as the number of relevant citations indicated by the user when she/he examines the search results, divided by the number of citations retrieved; (b) the user's satisfaction with the proportion of relevant citations retrieved; (c) the user's concern for recall in considering the search results; (d) the user's assessment of the value of the search in meeting

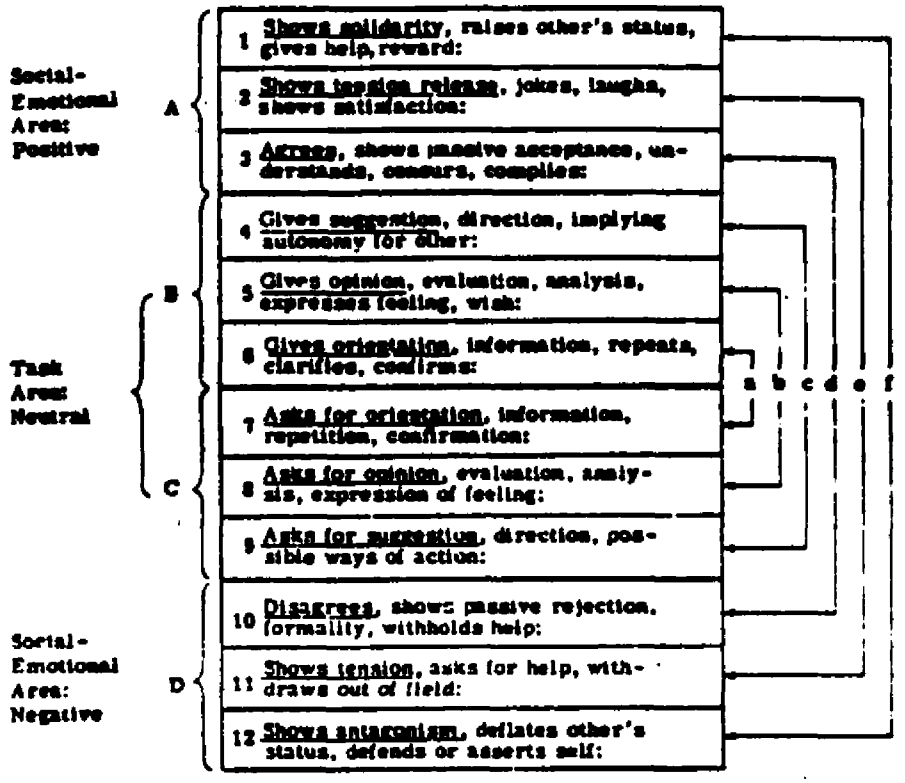
the need prompting the search request; and (e) the user's perception of the searcher in the interview process. The instrument can be considered to be valid insofar as it reflects common elements noted in other evaluation forms. Segments of the form which employ summative measures (concern for recall, perception of the interview) appear to be sufficently reliable when item consistency is considered. Reliability coefficients (alpha) were .76 for the recall concern score, and .90 for the search interview score (Nunnally, 1967, p. 196).

The major departure from most of the other evaluation forms reviewed, was the use of a 10-point rating scheme for many of the questions, in contrast to the three- or four-point scales most often used. Application of expanded scales is consistent with general scaling theory (Nunnally, 1967, p. 521).

Coded transcripts

Transcripts of selected interviews were unitized by the investigator and coded by trained judges. The categories applied to the transcripts were those defined by Bales in Interaction Process Analysis (1950). The scheme is outlined in Figure 1.

Bale's system is general in nature and applicable to small-group interaction. Although the system covers the gamut of interactive events that could occur in interviews or other interactive sequences, albeit in a very broad sense, the particular categories of interest in this study are those noted



KEY:

a Problems of Communication
b Problems of Evaluation
c Problems of Control
d Problems of Decision
e Problems of Tension Reduction
f Problems of Reintegration

A Positive Reactions
B Attempted Answers
C Questions
D Negative Reactions

Figure 1. Categories in the Interaction Process Analysis Scheme.

Note: Reprinted from Interaction Process Analysis by R. F. Bales, 1950, p. 9, by permission of the University of Chicago Press, Copyright 1949 by the University of Chicago.

as Task Area: Neutral in Figure 1, the information-giving and question-asking categories. The system appeared to be appropriate for this investigation because it is an established method of content analysis (Borgatta & Crowther, 1965; Heyns & Lippit, 1954, pp. 370-404; Holsti, 1969); definitions for scoring the categories were available (Bales, 1950, pp. 177-195); categories 4 through 9 were directly related to areas being investigated; "norms" in the sense of profiles for other small-group interactions existed (Bales & Hare, 1965); and it could be applicable to any future studies of search interviews.

Data Collection

In this section preliminary procedures, data collection period, site procedures, collection of evaluation forms, selection of tapes for transcription, and coding of transcribed tapes are described.

Preliminary Procedures

Prior to initiation of data collection a meeting was held to familiarize local participating searchers with the procedures and tape-recording process. Out-of-state searchers were mailed the procedures (Appendix A), forms, and blank tapes. At this same meeting comments were received concerning the user request form and the user evaluation form; some revisions were made based upon the input received. Searchers were aware that a study of the tape-recorded interviews constituted a part of the research plan, but they were not aware of the specific

methodology employed (i.e., the focus on information giving and question asking).

Data Collection Period

Site data was collected over a period of 4-1/2 months, from mid-June through the end of October 1978. Collection periods for each site were not concurrent, since final approval of the participant consent form (see Appendix A) occurred within different time frames. Two sites collected data for approximately 3-1/2 months, one site collected data for a lesser period.

User evaluation forms were accepted through mid-December 1978. The sequence of initial contact by mail, with two mail follow-ups and a telephoned reminder, required almost two months for completion after the end of site data collection.

Site Data Collection

Users requesting a MEDLINE search at the sites were given a consent form (Appendix A) which briefly outlined the project plan. If the user agreed to participate and signed the form tape-recording began. Interviews were then conducted in the usual manner.

Several searchers commented that they were quite aware of the recording process. However, since there was no preknowledge of what was being examined in the interviews, this does not appear to be an influencing factor in the ultimate consideration of the interviews. When the tapes were later reviewed by the investigator, no artificiality in interviewing procedures was readily apparent.

At the end of the interview the tape recorder was turned off. If on-line citations were given to the user the number of citations was noted. If the total number of off-line citations was known at the time, that number was also noted. In most cases the total number of citations was not immediately known, since backfiles were also searched. For these cases citation counts were recorded when the results were received.

The investigator periodically visited the two local sites to pick up tapes, consent forms, and request forms. The non-local site mailed materials on a weekly basis.

During the data collection period a total of 132 interviews were recorded at the three sites. Four of these interviews did not result in a MEDLINE search because after discussion with the searcher the user was directed to an alternative information source; two interviews were recorded but the request forms were not available.

Of the remaining 126 unique users, 65 were determined to be inappropriate to this study. Almost one-third of this number was excluded because the search results were not intended for the person present at the interview; thirty-five users were excluded because they indicated a status outside the scope of this study (e.g., attorneys, librarians, intern-residents). Four of the remaining users were excluded even though they were faculty because they indicated a non-research purpose in requesting the search; five were omitted because more than one person was interacting on the tape. The residual

pool of appropriate users included 23 faculty members and 38 students (N = 61).

Although the recording procedures led to recording of more interviews than would ultimately be utilized, this seemed appropriate. Interaction which would eliminate users at the outset (e.g., determination of status, purpose, etc.) is considered in Taylor's five-filter model (1968) to be an integral part of the interview process.

User Evaluations

All appropriate users were sent a cover letter (Appendix A) explaining the project and an evaluation form. A return envelope was enclosed. If the evaluation form was not returned after 2-1/2 weeks, a hand-written note was sent to the user prompting her/him to return the form. If, after another two weeks, the form was still missing a third note was sent. As a final step an attempt was made to contact the user by phone and urge the return of the form.

A total of 55 forms were returned (90%); however, one of the forms was returned two months after the cut-off date and was not included in the analysis. For the 54 included searches, 22 were from faculty (96% return) and 32 were student searches (84% return).

Transcription of Tapes

Although it was originally intended that interviews for transcription would be randomly selected from the pool of cases with completed evaluation forms and good recordings, this did not prove feasible. Preliminary work with the tapes indicated

that they would be more difficult to transcribe than had been anticipated from an earlier report of similar transcription (Carmon, 1975). A more recent study involving taping of interviews in public libraries (Lynch, 1978) suggests that transcription is a lengthy, complex process. In the present study some familiarity with MEDLINE procedures and medical terminology proved to be necessary for an accurate transcription of the tapes. For this reason the investigator transcribed all tapes rather than delegating them to clerical personnel as originally intended.

Even though the investigator was more familiar with the context and terminology involved, transcription proved to be a time-consuming task, involving in several instances up to 20 hours per interview. This was particularly true for those interviews which involved use of the terminal while the user was present. Dialogue had to be detected against the background noise of the operating terminal. Knowledge of the length of time likely to be involved in transcribing each interview indicated that a delay for random selection subsequent to the return of all evaluation forms was not appropriate within the time-frame of this study.

Interviews were selected for transcription by two criteria: first, to provide diversity in the searchers and user types included, and second, to allow transcription to occur along an extended period. Interviews thus selected were those associated with the first faculty and first

student evaluation forms returned for each searcher. This selection does not necessarily reflect the initial interviews recorded by searchers. Some searchers had recorded several inappropriate users before encountering a request which met study specifications. Other searchers recorded an appropriate user early in the process, and were involved with an alternate appropriate user much later on. Two searchers had only student interviews available for transcription. A total of 18 interviews were transcribed (10 students, 8 faculty). One searcher had conducted relatively few interviews, none were appropriate for transcription.

The interviews were divided into interaction units according to Bale's methodology (1950, p. 37) at the time of transcription.

Coding of Interviews

Transcribed interviews were coded by two experienced searchers at the NASIC Search Service, MIT Libraries. Coders are familiar with MEDLINE and other on-line data bases, and as full-time searchers are involved with interviewing on a daily basis. It was believed that this familiarity with the interview process would provide a good background for interpreting the unitized transcripts in accordance with the scoring system.

The coders were first sent copies of the definitions of categories (Bales, 1950, pp. 177-195) and copies of several transcripts obtained from the University of Georgia study

(Carmon, 1975), so that they might become familiar with the coding scheme. Once the coders had some practice at applying the scheme to interviews an extended training session was scheduled. A number of supplemental coding conventions were agreed upon at that time (see Appendix B). At the end of the training session coders were able to demonstrate 82% agreement when scoring three pages containing a total of 96 units (categories 4, 5, and 6 were collapsed, as were categories 7, 8, and 9 since data manipulation would involve total question-asking by the searcher and total information-giving by the user).

Transcribed interviews for this study were divided between the coders so that each received an approximately equal number of pages to score. A duplicate copy of one search was sent to each coder so that a reliability coefficient for coding material involved in this study could be established. Three pages were selected at random from the search for the determination. A more stringent method (Cohen, 1960) than simple percent agreement was applied (see Appendix B). A coefficient of .77 was obtained; this appeared adequate for the applications involved in this study which focused on the ranking of results.

Interview scores for the analyzed searches were obtained by counting the frequency of events coded by category for the searcher and the user. Topics of the 18 transcribed and coded searches are listed in Appendix C. Appendix D includes an example of a coded interview.

Data Treatment

Distribution-free tests appeared to be most appropriate for analyzing results. This decision was based upon two considerations. First, there is a considerable body of evidence which suggests that data resulting from bibliometric studies are not normally distributed (Saracevic, 1970a). Second, although scale values are used to assess user responses on the evaluation form, the instrument has not had a wide application, and it appeared more prudent to reject assumptions of interval level data.

Specifically, in this study the first hypothesis (the interrelationship of the value assessment and relevance, satisfaction with the proportion of relevant items retrieved, concern for recall, and the perception of the search interview) is examined by means of the Kendall rank correlation coefficient. The second through seventh hypotheses (differences in user groups) are examined by the Mann-Whitney U test. Hypotheses eight through ten (relationship of information-giving by the user, to question-asking by the searcher, and assessments of relevance and concern for recall) are also examined by Kendall's rank correlation. Finally, hypotheses eleven and twelve (searcher question-asking activity as a contributory factor in user assessments of relevance and concern for recall) are examined by means of the Kendall partial rank correlation coefficient. Version

7 of the SPSS program was utilized for the computations (Nie, Hull, Jenkins, Steinbrenner, & Brent, 1975).

The following chapter describes the results of this study.

Chapter 4

Data Analysis

Data resulting from this study is presented and analyzed in four sections of this chapter. The first section describes the data sources. In the second section the 12 research hypotheses from Chapter 1 are tested. The third section discusses the results of these tests and interprets them in light of other findings related to the scope of this study. The last section examines the search interview and profiles developed in the coding process; and discusses the profiles in relation to this study and to other "norms" reported for interaction profiling.

Data Sources

General variables considered are available from the user request forms and the evaluation forms corresponding to the 54 users who completed the evaluation aspect of this study. This group included 22 faculty users and 32 student users. One of the student users did not indicate the number of relevant citations retrieved on the returned form, thus the number of cases for any testing which includes relevance is 53.

Interviews from 18 of the 54 cases were transcribed and analyzed. This subset includes 10 student users interacting with 10 different searchers, and 8 faculty users interacting with 8 different searchers. Since the subset does not

represent a random selection from the entire group of evaluating users, the evaluative responses of the two groups (coded interview, non-coded interview) were analyzed by means of the Mann Whitney U statistic. Analysis ($p = .05$) indicates that there is no difference in the number of relevant citations retrieved, the relevance scores, the satisfaction with the proportion of relevant citations, the concern for recall scores, the value scores, or the search interview scores indicated by these two groups. Thus, it appears safe to assume that the searches selected for coding are valid representations of the entire sample studied, insofar as evaluative measures are concerned.

Hypotheses Tests

The 12 research hypotheses defined in Chapter 1 are examined in three sections. The first section examines the hypothesis relating to interrelationships of search assessment measures and the overall value designation for the search. The second section examines user differences in assessment and interview activity. The last section examines specific interview activity and assessment measures.

Interrelationships in Evaluation

The 54 cases (or 53 where noted) are examined for relationships in evaluation measures.

H_1: User assessment of search value is not related to user assessment of the relevance score, the satisfaction with

the proportion of relevant citations score, the concern for recall score, or the search interview score.

When the evaluation scores are ranked and compared to the ranked scores for value by means of the Kendall rank correlation coefficient, moderate, but significant correlations are noted in each instance. Table I indicates, in decreasing order of relationship, the tau values observed for the cases.

TABLE I

Correlation of Evaluative Measures
With User Designations of Value

Measure	Kendall's tau
Satisfaction with the proportion of relevant citations retrieved	.66*
Concern for recall	.60*
Relevance[a]	.47*
Search interview	.42*

[a]N = 53

*p = .001

User Differences

Hypotheses 2 through 4 suggest that faculty users will assign lower relevance scores, concern for recall scores (i.e., by the definition be more concerned about recall) and search interview scores, than will student users. Hypotheses 5 and 6 suggest no differences in faculty and student scores

for satisfaction with the proportion of relevant citations retrieved and value. Hypothesis 7 suggests greater information-giving activity by faculty users in the coded interviews. These hypotheses are tested by means of the Mann-Whitney U statistic.

H_2: There is no difference in the distribution of faculty and student relevance scores.

Results do not allow rejection of the null hypothesis.

H_3: There is no difference in the distribution of faculty and student recall concern scores.

Results are not in the direction predicted, therefore the null hypothesis is not rejected.

H_4: There is no difference in the distribution of faculty and student search interview scores.

Results do not allow rejection of the null hypothesis.

Table II is a summary of the results from testing the second through fourth research hypotheses.

TABLE II

Summary Results for Testing User Differences on Relevance, Concern for Recall, and the Search Interview

Measure	Mean Rank by Group Faculty (N=22)	Student (N=32)	Mann-Whitney U	Z	P (one-tailed)
Relevance	28.00	26.29[a]	319.0	.397	.35
Concern for Recall	32.39	24.14	244.5	1.895	.03[b]
Search Interview	27.64	27.41	349.0	.053	.46

[a]31 students.

[b]Direction opposite from that predicted.

H_5: <u>There is no difference in the distribution of faculty and student satisfaction with the proportion relevant scores</u>.

Results do not allow rejection of the null hypothesis.

H_6: <u>There is no difference in the distribution of faculty and student value scores</u>.

Results do not allow rejection of the null hypothesis. Table III outlines the results from testing H_5 and H_6.

TABLE III

Summary Results for Testing User Differences on Satisfaction with the Proportion of Relevant Citations and Value

Measure	Mean Rank by Group: Faculty (N=22)	Mean Rank by Group: Student (N=32)	Mann-Whitney U	Z	P (two-tailed)
Satisfaction with proportion relevant	31.52	24.73	263.5	1.577	.12
Value	30.52	25.42	285.5	1.187	.24

H_7: <u>There is no difference in the distribution of faculty and student information-giving activity</u>.

Results do not allow rejection of the null hypothesis. Table IV indicates results from this testing.

TABLE IV

Summary Results for Testing User Differences in Information-Giving

Measure	Mean Rank by Group: Faculty (N=8)	Mean Rank by Group: Student (N=10)	Mann-Whitney U	Z	P (one-tailed)
User Information-giving	8.31	10.45	30.5	.845	.20

Interview Activity

Hypotheses 8 through 12 examine interrelationships of user information-giving activity, searcher question-asking activity and user assessments of relevance and concern for recall. The hypotheses are examined by means of the Kendall rank correlation coefficient and Kendall's partial rank correlation coefficient.

H_8: In the search interview information-giving activity by the user is not related to the relevance score.

When user information-giving activity scores (the sum of user activity coded for categories 4, 5, and 6 in Bales' scheme) are ranked and compared to ranked relevance scores, a moderate but significant relationship is noted.

H_9: In the search interview information-giving activity by the user is not related to the concern for recall score.

Results do not allow rejection of the null hypothesis.

H_{10}: In the search interview question-asking activity by the searcher is not related to information-giving activity by the user.

When searcher question-asking scores (the sum of searcher activity coded for categories 7, 8, and 9 in Bales' scheme) are ranked and compared with user information-giving scores (as previously defined), a moderate, significant relationship is observed. Table V summarizes correlation data from hypotheses 8 through 10.

TABLE V

Correlation of User Information-Giving With Relevance, Concern for Recall and Searcher Question-asking

Measure	Kendall's tau
Relevance	.50*
Concern for recall	.20
Searcher question-asking	.58**

Note: Data from 18 coded interviews

*p = .004
**p = .001

H_{11}: There is no relationship between user information-giving and relevance scores when searcher question-asking is held constant.

Kendall's partial rank correlation is utilized to examine the relationship that remains between information-giving activity and relevance when the effects of variation in question-asking activity are eliminated. The relationship is determined by the partial rank correlation formula (Siegel, 1956, p. 226).

$$\tau_{xy.z} = \frac{\tau_{xy} - \tau_{zy}\,\tau_{xz}}{\sqrt{(1 - \tau_{zy}^2)\,(1 - \tau_{zx}^2)}}$$

where $\tau_{xy.z}$ = the partial rank correlation coefficient

τ_{xy} = the rank correlation coefficient for information-giving and relevance

τ_{zy} = the rank correlation coefficient for question-asking and relevance

τ_{xz} = the rank correlation coefficient for information-giving and question-asking

τ_{xy} (.50) and τ_{xz} (.58) were previously determined. τ_{zy} is determined as .36 (p = .04). Using these values a partial rank correlation coefficient ($\tau_{xy.z}$) of .38 is noted.

From observation an enhancing factor for question-asking is suggested, since the initial relationship between information-giving and relevance is reduced somewhat when the effects of question-asking are eliminated. However, there are currently no methods for testing the significance of the partial coefficient obtained (Conover, 1971, p. 254).

H_{12}: <u>There is no relationship between user information-giving and the recall concern score when searcher question-asking is held constant</u>.

This relationship was not examined because the earlier test (H_9) indicated no significant relationship between user information-giving activity and the concern for recall score.

Discussion of Results

User Evaluations

Like users in most other reports of on-line evaluations, MEDLINE users, when considering the value of the search results in meeting the need prompting their request, indicated generally favorable responses concerning value. This was true even though users were provided a ten-point scale to indicate value rather than the more customary three- or four-point scale. No users marked the lowest designation (no value) and almost 60% of the users indicated a value of eight or higher. Figure 2 outlines the distribution of responses on the value scale. Application of an extended scale does, however, appear to have some merit in increasing the variation of response patterns, which is particularly useful if comparisons of magnitude, as in this study, are desired.

The favorable response pattern on value is coupled with a fairly low pattern of relevance scores. The mean relevance score for all searches is 45%. Thirty-one of the searches had relevance scores of 50% or less (i.e., in these searches half or fewer of the citations sent to the user were considered to be relevant to the search question). Figure 3 outlines the distribution of relevance scores. The distribution tends to be U-shaped, a pattern which was noted by Rees and Schultz (1967, Vol. 1, p. 118) when they pooled relevance ratings for all documents in their experimental study of relevance.

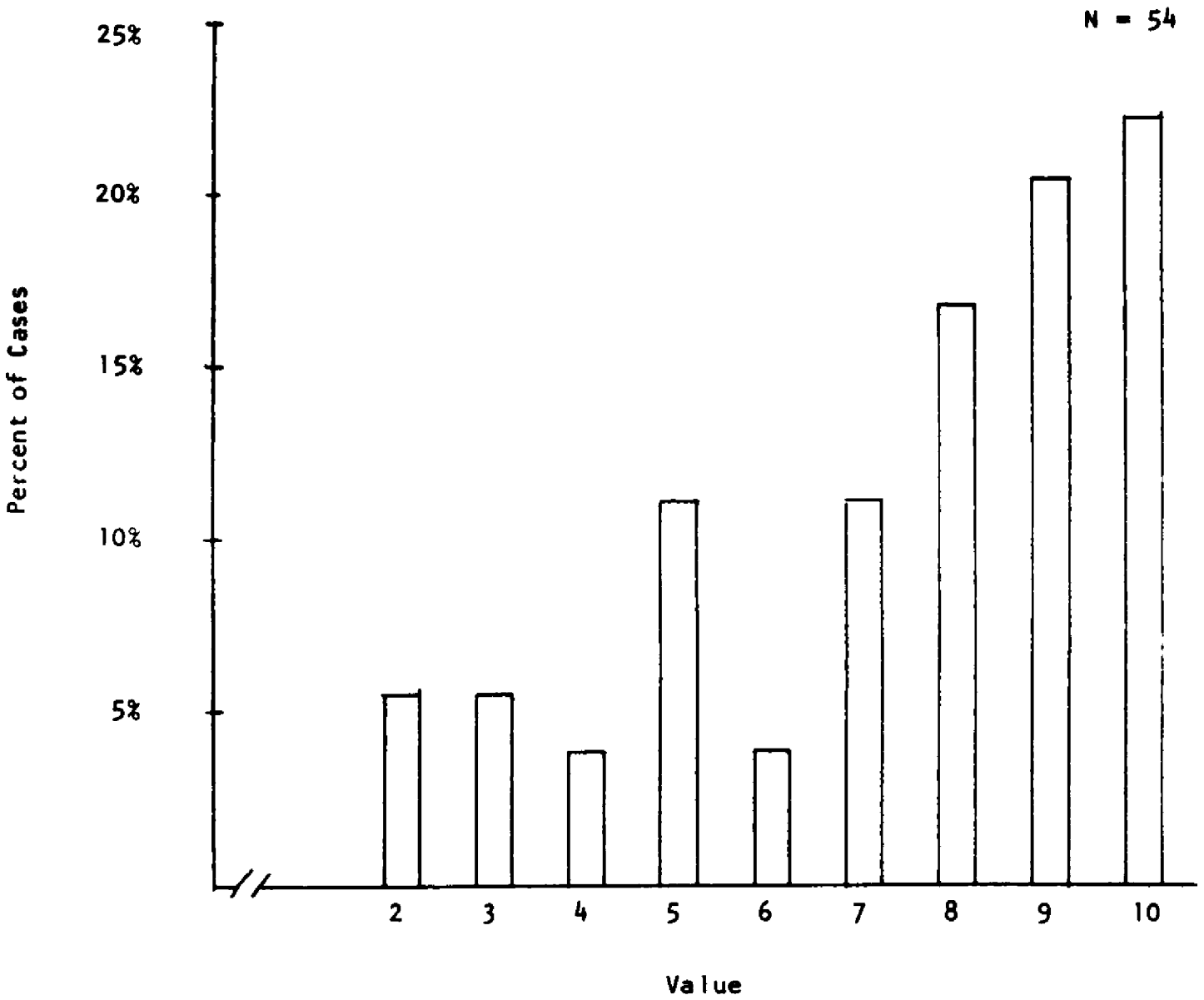


Figure 2. Value of MEDLINE Search Results in Meeting the Need Prompting the Search Request.

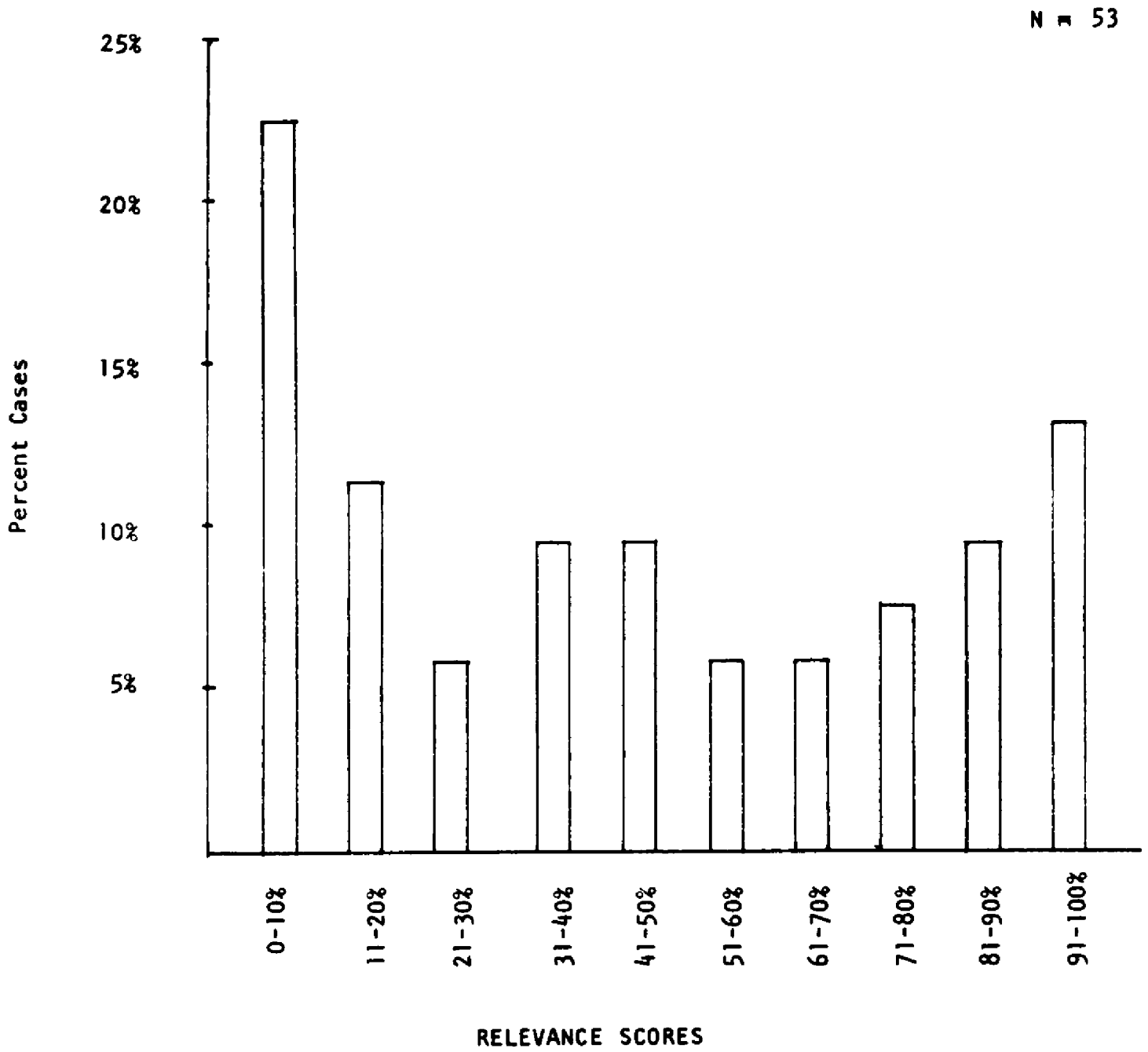


Figure 3. Distribution of Relevance Scores for MEDLINE Searches.

As noted previously, relevance and value scores show only a moderate relationship to each other (τ = .47). Tagliacozzo (1977) has suggested that dissonance in user responses may occur; in that study dissonance (contrary responses) were noted for helpfulness and usefulness. In the present study dissonant responses appear to be concentrated in cases with lower relevance scores. For example, all cases of medium (5 to 7) to high (8-10) value designations are associated with medium to high relevance scores, but lower relevance scores (0-33%) are associated with low, medium, and high value designations. Relevance scores are simply ratios determined by dividing the number of relevant citations by the total number of citations retrieved, and do not reflect the actual number of "good" citations the user gets from the search. For example, a user might find four relevant citations out of a total of five retrieved; the relevance score is 80%. Another user might find 40 relevant citations in a search which results in a total of 80 citations; the relevance score in this case is only 50%. The second user, however, has more relevant citations. It has been suggested that searches that produce more relevant references are associated with helpfulness (Tagliacozzo, 1977, p. 246) or satisfaction (Kobelski & Trumbore, 1978, p. 16). In the first case searches with six or more useful references were contrasted with those achieving five or fewer on a dichotomized helpfulness scale. In the second case the authors noted

that all users receiving 40 or more relevant citations were satisfied. They also indicate that all users receiving total outputs of 100 or more citations were satisfied.

In the present study 68% of the users received 40 or fewer relevant citations. Figure 4 indicates the distribution of relevant citations. By sheer volume of relevant citations, the results in the current study appear to be "better" than those indicated in the public library experiment (Mick, 1977). In that study 60% of the second year users received 0 to 15 relevant citations, in the third year 75% of the users received 0 to 15 relevant citations. Yet it can be recalled that 68% to 76% of the users indicated that their search was of major or considerable value. This suggests that the actual number of relevant citations received by the user may not be of paramount importance in value estimations. In the current study of MEDLINE users, when the number of relevant citations received are ranked and compared to ranked user value designations a moderate relationship (τ = .53) similar to that noted for relevance is observed. A negligible relationship (τ = .16) appears to exist between the total number of citations retrieved and value.

The obtained search interview score (the mean score computed for five questions to the user about the searcher's role in the interview) showed the least relationship to the value score when the hypotheses measures were tested (τ = .42).

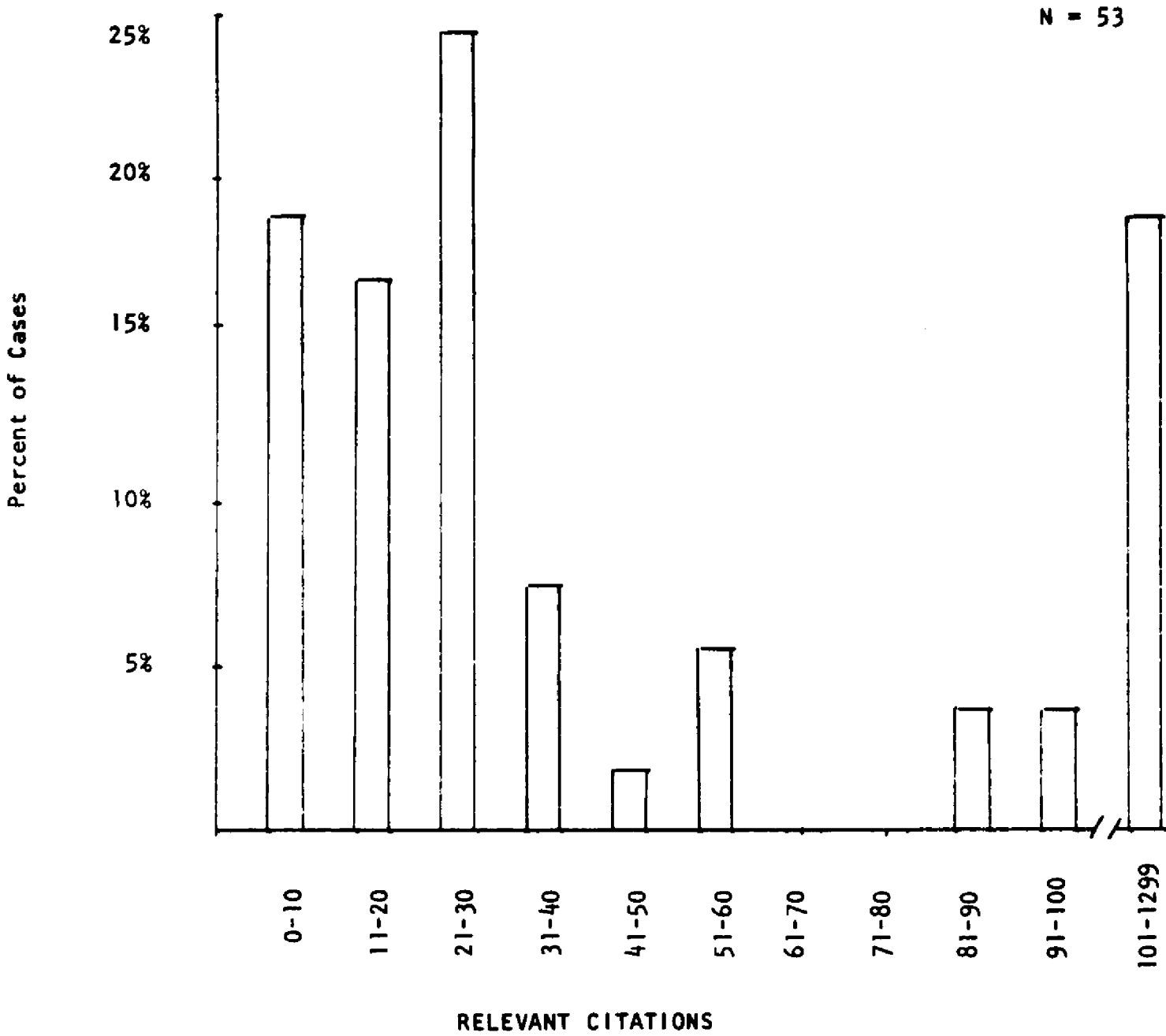


Figure 4. Distribution of Relevant Citations in the MEDLINE Searches.

User impressions of the searcher in the interview situation were generally favorable. Scores of 7 or higher were observed for 77% of the searches. Figure 5 indicates search interview scores for the MEDLINE searchers. Findings reported here do not appear to be inconsistent with those reported earlier by Gothberg (1975) concerning user perceptions about two types of interview situations and user satisfaction with the transfer of information. Users exposed to immediate interviews (a sense of "liking" is conveyed) were more satisfied with the interview itself and better satisfied with their own performance in the interview than were users exposed to non-immediate interviews. However, the two user groups exhibited no difference in their expressed satisfaction with the transfer of information. Although the conditions are different, one might anticipate from the earlier study that the user perception of the interview (a social interaction) will exhibit a somewhat lesser relationship with a judgment of value than other areas more closely aligned with the product (the search results) being evaluated.

Most closely associated with the value designations indicated by the user were the scores on satisfaction with the proportion of relevant citations retrieved (τ = .66) and the mean score concern for recall (τ = .60). Some of the dissonance noted at the bottom third of the relevance scale (high, medium and low value scores) may be related to

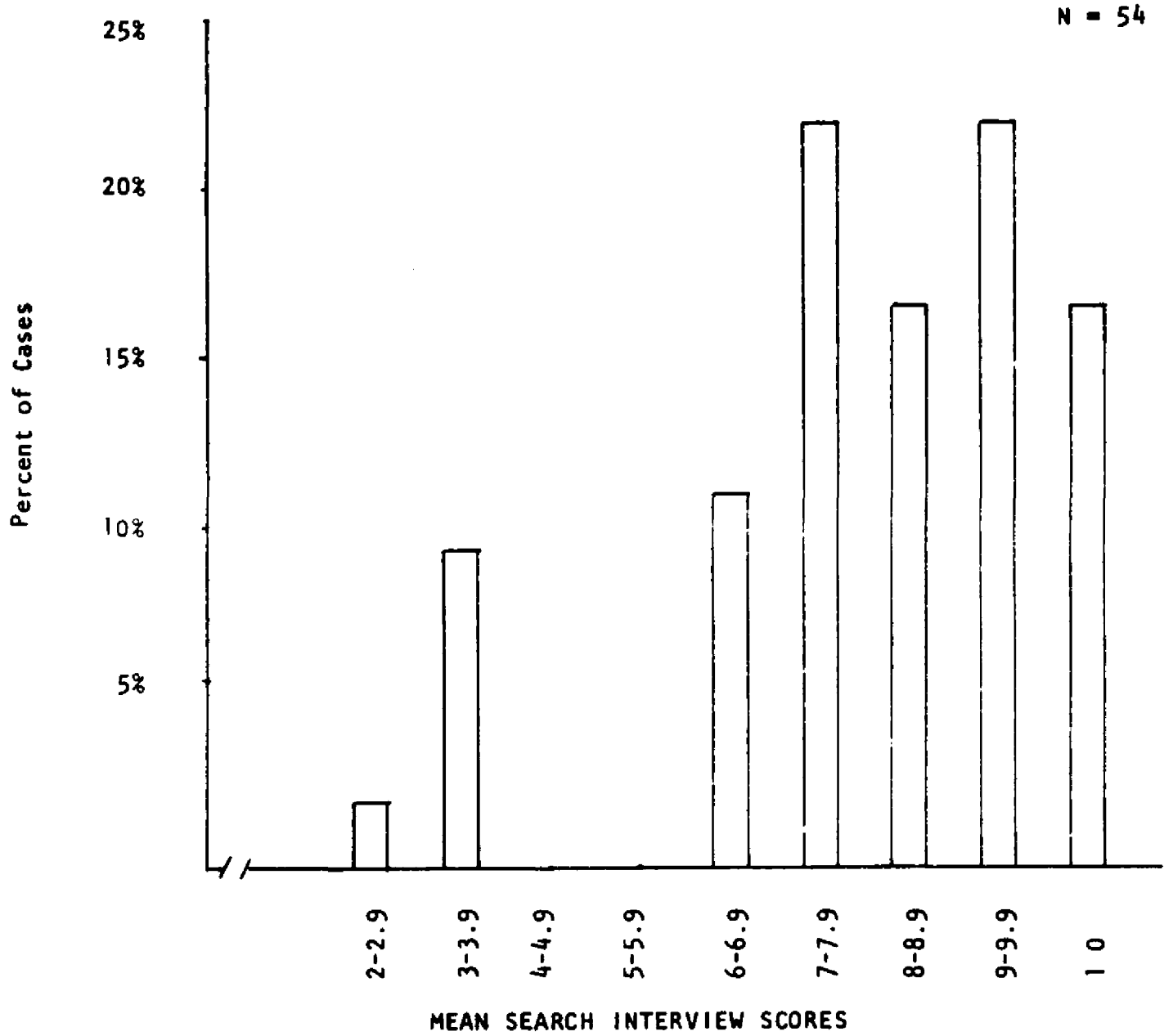


Figure 5. Distribution of Mean Scores for User Perceptions of the Search Interview.

a perception by some users that lower relevance means that they have missed some items. Eight of the nine users indicating a satisfaction with the proportion relevant score of three or less had relevance scores lower than 30%. Five of these eight had concern for recall scores less than three.

In the Georgia-UCLA study (Carmon, 1975), users ranked the lack of ability to judge completeness as the most disliked feature of the search services. In the current study the concern for recall score exhibits a considerable relationship with the value score. Repetition in this study of a situation in which users appear to attach some importance to the recall aspect of a search is particularly interesting in light of Cooper's suggestion (1973) that recall may be an inappropriate measure for consideration in retrieval evaluations, i.e., he does not consider it important to the user because, in a general sense, the user cannot "know" that he has missed citations in the results. Figure 6 indicates concern for recall scores. Figure 7 indicates the satisfaction with the proportion of relevant citations.

User Differences

Earlier reports (Rees and Schultz, 1967; Saracevic, 1970a) provided a basis for belief that, in relevance assessments, knowledgeability was associated with more stringent assessment. In the current study these findings were extrapolated to a model which suggested that this same stringency of assessment would occur in actual searches of

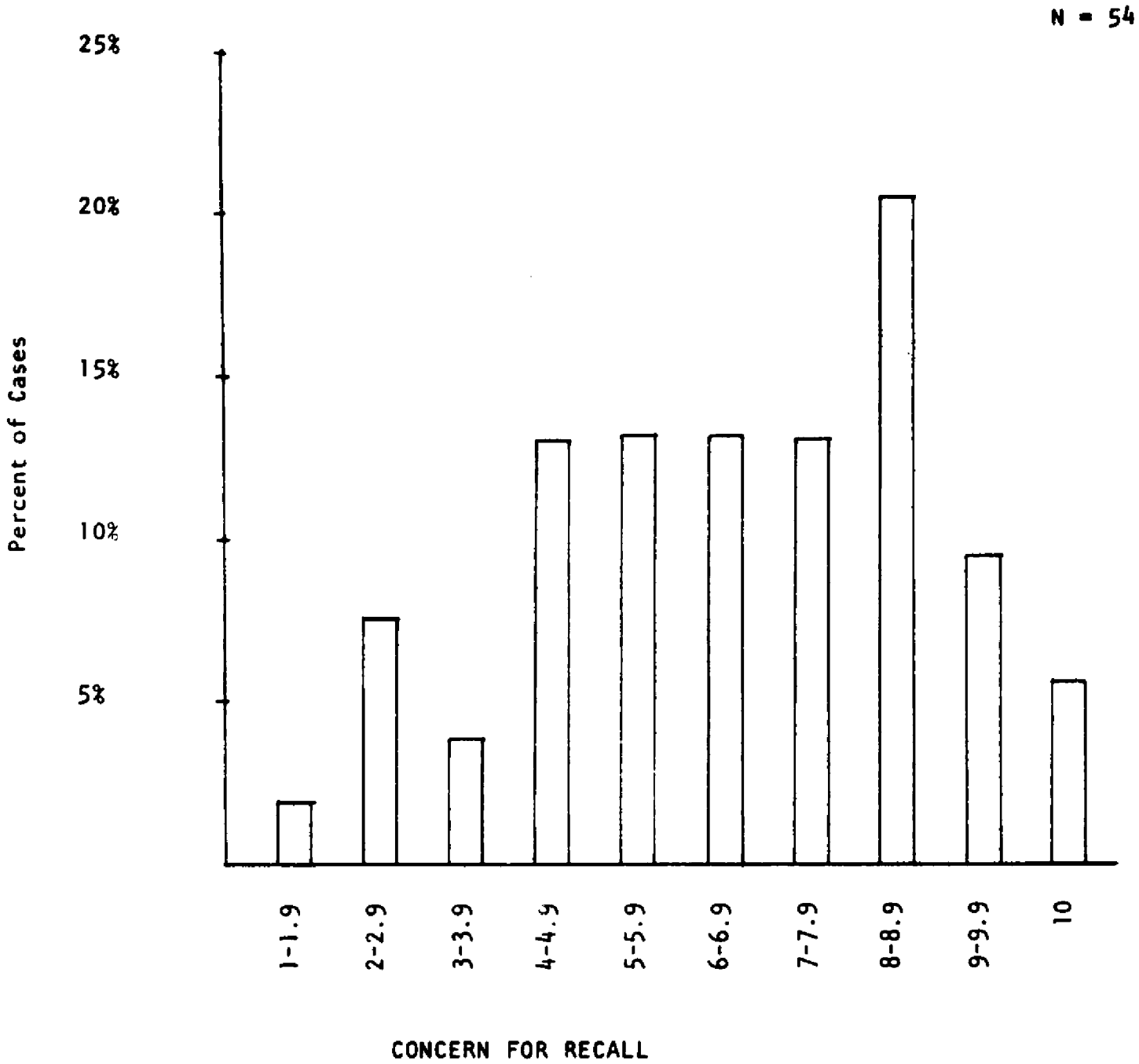


Figure 6. Distribution of Concern for Recall Scores.

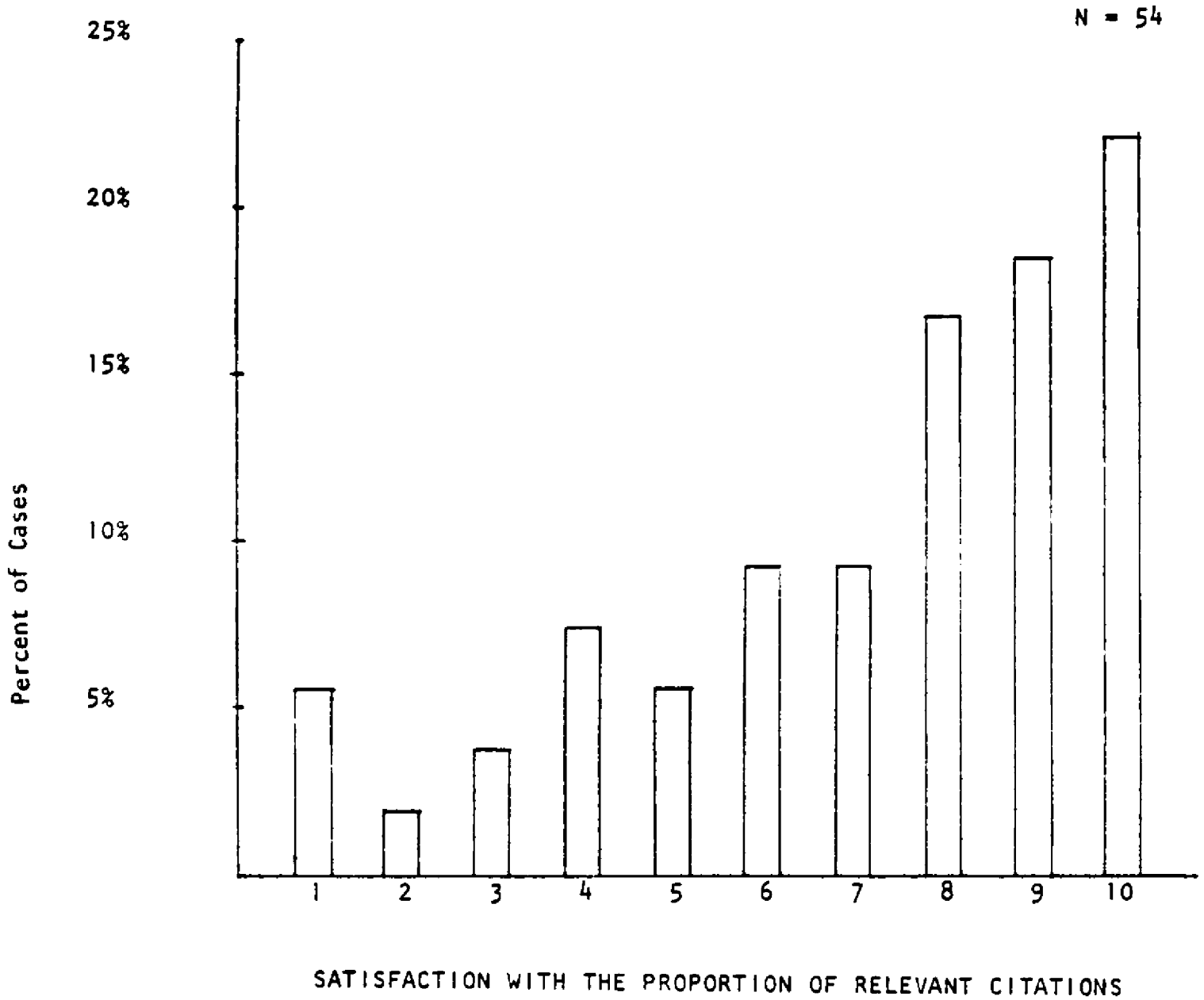


Figure 7. Distribution of Satisfaction with the Proportion Relevant Scores.

an on-line data base. Specifically, it was suggested that a more knowledgeable group of biomedical users (faculty) would indicate lower relevance scores, lower concern for recall scores (be more concerned about recall), and lower search interview scores than would a less knowledgeable group of users (students). Since value and satisfaction with the proportion of relevant citations retrieved had no immediately apparent connection with knowledgeability, it was suggested that the two groups would exhibit no differences in these areas. When the interview itself was considered it was predicted that the more knowledgeable users would exhibit greater information-giving activity.

As anticipated no difference in user groups was noted on the two more general measures of value and satisfaction with the proportion of relevant citations. However, in areas where it was expected that faculty users would make more stringent assessments, this did not occur. Faculty and student users exhibited no significant rank order differences in the direction predicted for relevance scores, recall scores, or search interview scores. In information-giving activity faculty users exhibited no significant rank order differences when compared with students.

Since the expectation of more stringent assessments and more information-giving activity from faculty was based on the supposition that faculty users were more knowledgeable, several items from the user request form were examined to

determine if indicators of knowledgeability were present in the sample studied.

Utilizing responses from question 12 on the request form, degrees completed by faculty and student users were examined. To the extent that knowledgeability is equated with higher levels of educational attainment, faculty users can be considered more knowledgeable when the proportion of users in each category is considered (Kendall's tau C = .94, $p = < .001$). Eighteen students had completed one degree, the Bachelor's; eleven students had an undergraduate degree plus the Master's. These groupings were descriptive for none of the faculty. Ten users had completed a professional degree program (M.D., J.D., D.D.S.) or had two Master's degrees. Three students were in this category. Fifteen users had a Ph.D. or a Ph.D. plus another degree (e.g., M.D.). No students were included in this category.

Responses to question 13 on the request form were examined for variables related to general knowledge of, and continuing interest in, the topic of the search. No significant association is noted between user status and the awareness of several recent publications on the topic of the request (19 out of 21 faculty responded affirmatively, as did 23 out of 32 students), or between status and the user's continuing interest in the topic (all faculty indicated continuing interest, 30 out of 31 students indicated the same interest). However, faculty users were more likely to

read on a regular basis journals on the topic of the request ($X^2 = 4.19$, $p = .04$), and to have published an article on the topic ($X^2 = 4.31$, $p = .04$).

By topic the evaluated searches can be subjectively divided into two areas, conventional biomedical searches and social science searches (e.g., spirtuality of patients, sexuality and the mentally retarded). If mean rank scores by area of search (41 biomedical searches, 13 social science searches) are examined on relevance, satisfaction with the proportion of relevant citations retrieved, concern for recall, value, and the search interview, it is noted that on all measures biomedical related searches rank somewhat higher. It might be thought that area of the search affected the results obtained. However, there is no apparent association between user groups and type of search initiated.

Faculty users were more likely to have participated in a previous on-line interview than were student users (77% of the faculty had previous experience in contrast to 38% of the students), yet if cases are divided by first-time and previously experienced users, mean rank scores on the assessment parameters do not parallel those noted earlier (Tables II-IV) for the faculty and student user groups.

On all assessment parameters for cases in which the terminal session was included with the interview higher mean ranks for terminal sessions are noted. However, there

was no significant association between user groups and type of interview.

From a broader perspective it appears that one could reasonably anticipate that user groups like those defined in this study will exhibit no differences in assessment parameters. There appears to be little reason to believe that aspects of the particular cases (area of the search, previous experience, type of interview session) were related to the absence of more stringent assessments by faculty users.

Search Interviews

The relationship of user information-giving activity during a search interview to traditional areas of evaluation concern (relevance and recall) was examined. User information-giving scores for 18 selected interviews were ranked and compared to scores on relevance and concern for recall. Some relationship between ranked scores for information-giving and relevance was observed (τ = .50). The relationship between information-giving and the concern for recall score was negligible and non-significant. Information-giving was also related to question-asking activity by the searcher (τ = .58).

In Chapter 1 it was suggested that information-giving by the user can be either voluntary or searcher-elicited. To examine the remaining relationship between information-giving and relevance when the relationship to searcher

question-asking is eliminated, the partial rank correlation coefficient for information-giving and relevance was determined (τ = .38). The decrease in the observed relationship provides some basis for suggesting that, in general, question-asking by the searcher enhances the relationship between user information-giving scores and relevance scores.

The volume of information-giving units and question-asking units varied considerably over the evaluated interviews. The mean score for searcher question-asking units was 37.5 (S.D. 27.3); the mean score for information-giving units was 80.8 (S.D. 59.6).

Seven of the eighteen coded interviews included terminal sessions. More question-asking was noted for these interviews ($\bar{x}$ = 62.7, S.D. 26.3) than for non-terminal interviews ($\bar{x}$ = 21.5, S.D. 11.1). More information-giving was noted for users in terminal searches ($\bar{x}$ = 119.6, S.D. 51.6) than in non-terminal searches ($\bar{x}$ = 56, S.D. 52.1).

Although relevance itself has a moderate relationship with the value of the search results (τ = .47) and information-giving appears to be more associated with relevance (τ = .50) than does question-asking (τ = .36), some interactive effect appears to occur, which suggests that training in question-asking may be a factor for consideration in preparation of on-line searchers. To the extent that question-asking is important, the impetus, or opportunity for additional questions appears more likely

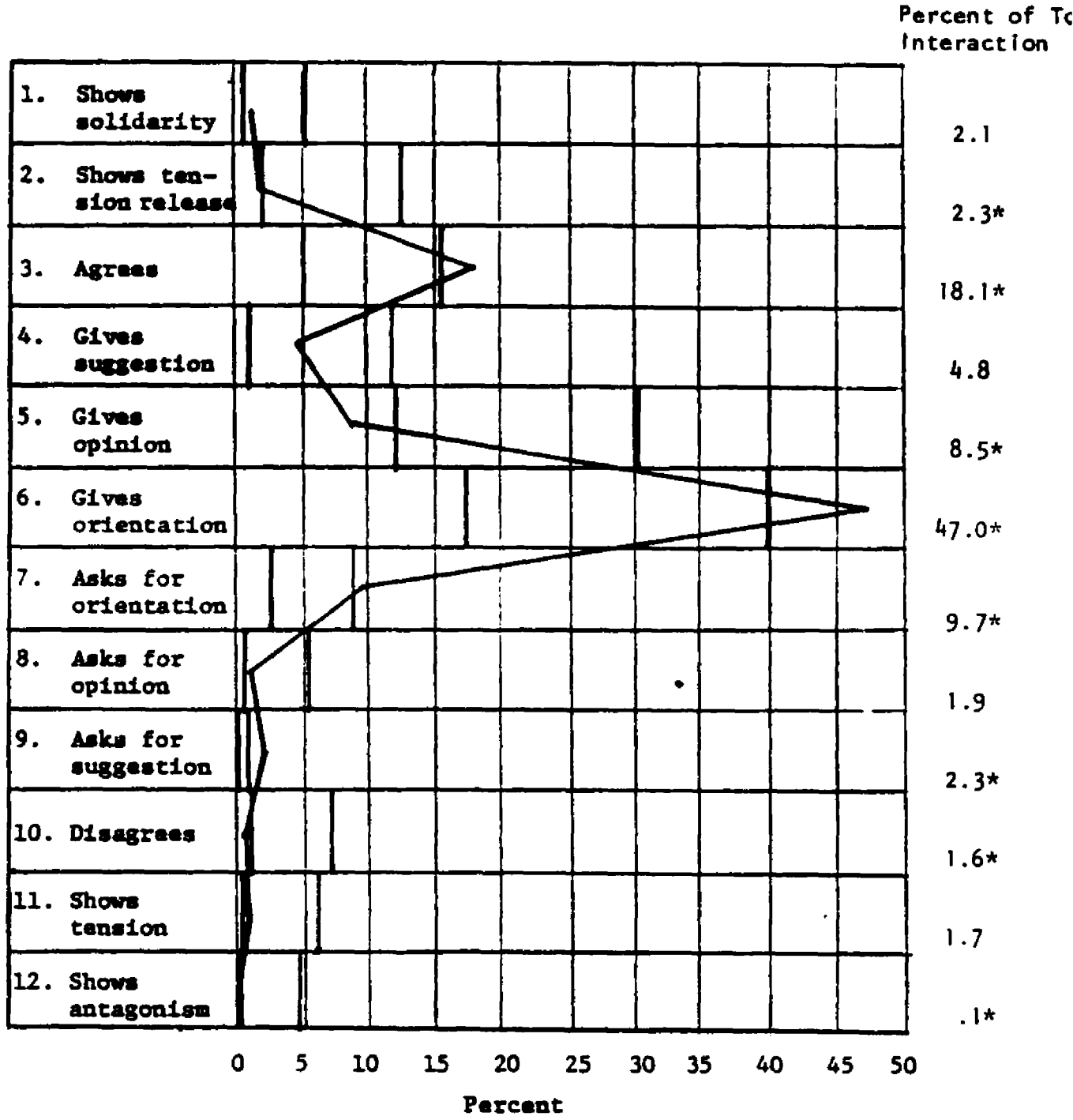
to arise in interactive searches at the terminal with the user present, than in interviews removed from the actual terminal search session.

Interview Profiles

In this section an overall profile for all coded searches is presented and compared with a standard summary profile which represents pooled findings from 21 studies which employed the Bales' scheme (Bales & Hare, 1965). Breakouts for searcher and user interaction, and interaction by type of interview (non-terminal, terminal) are also discussed.

Figure 8 gives the percentage of interaction by category for the MEDLINE interviews. Darker vertical grids indicate points 1 S.D. above or below those noted by Bales and Hare in the pooled interaction studies.

The most obvious departure occurs in the information-giving category. Agreement units and asking for suggestion units are also higher (+ 1 S.D.) than those reported previously. This pattern would appear to be consistent with the aim of the search interview "to arrive at a clear, narrative, natural language statement of the user's information needs and to gather a number of facts and clues to be used to amplify or refine this statement" (Carmon, 1975, p. 4). The profile is generally lower in those areas which indicate positive and negative emotional



Notes:

* = Percentage is above or below 1S.D. of diagnostic population.

Total number of interaction units = 6558

Number of searches = 18

Vertical grid marks indicate 1S.D. for the category.

Figure 8. Percent of Total Interaction by Category for 18 Transcribed MEDLINE Interviews.

areas. Some of this may be attributed to the method of coding, i.e., utilization of a typed-transcript. Waxler and Mishler (1966) noted lesser use of negative categories in typed transcripts (loss of emotional tone in interchanges) but found that an overall comparison (rank order) of category use for the two methods (typed-transcript, tape and typed-transcript) gave similar distributions (rho = .957, $\rho =$ .001). The lower rates noted for categories 1 and 2 may simply reflect the business-like nature of the interview situation. The higher percent of units scored for the agreement category (Category 3) parallels findings in other types of discussion groups where participants are required to reach consensus on an issue.

Figure 9 provides more illumination on who does what during the interview. It is apparent that the searcher dominates the task area of the profile (Categories 4 through 9). The information-giving activity by the searcher is most striking, since it seems that this giving of information is not necessarily the response to a high rate of question-asking activity by users. Users appear to take a relatively passive role in asking questions. If one of the aims of the interview is to reach a concensus on the needs of the user, it seems that the concensus that is reached may be somewhat influenced by the searcher. The searchers indicate higher levels of giving suggestion and opinion, while users are predominant in the agreement category.

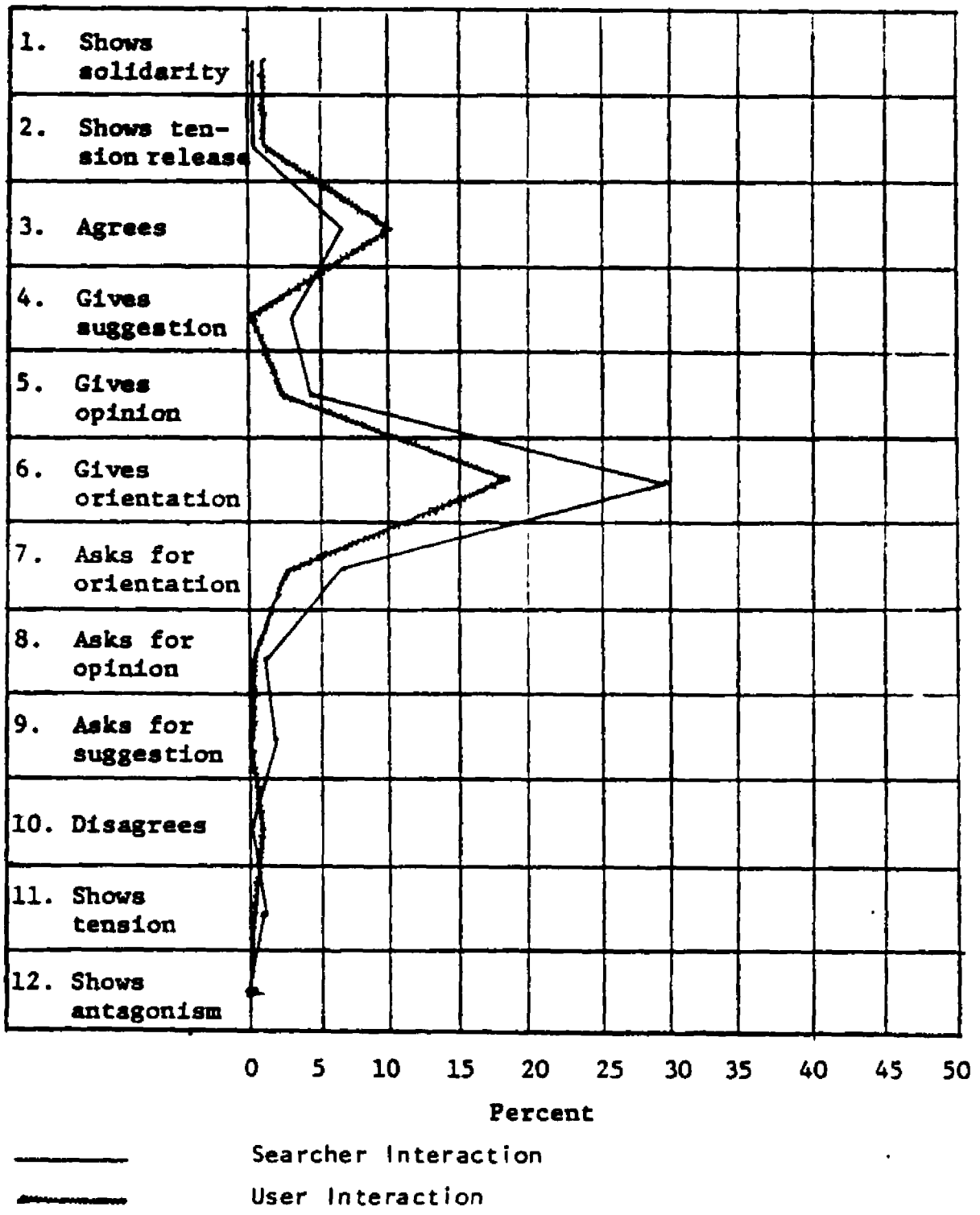
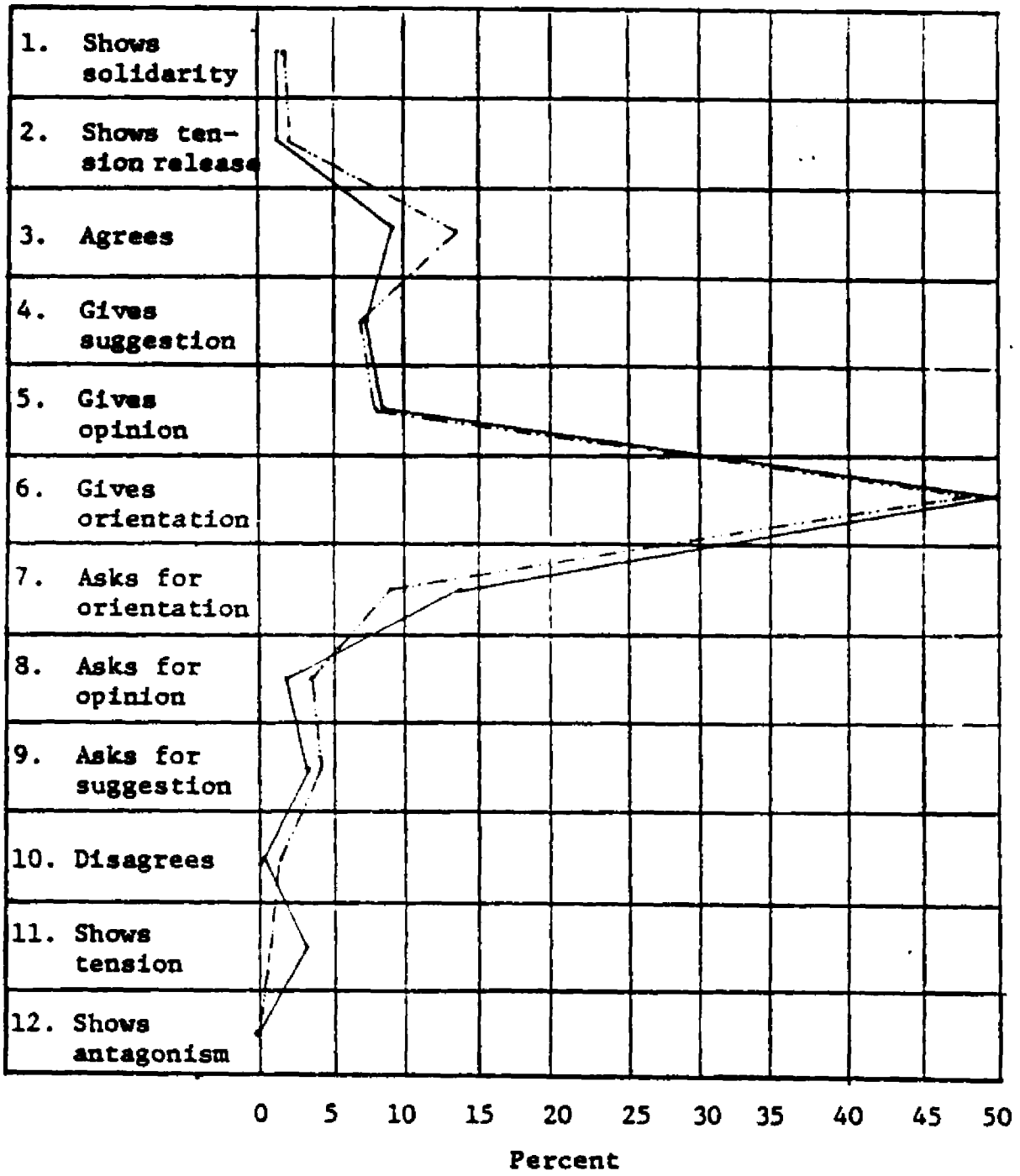


Figure 9. Percent Searcher and User Interaction by Categories for 18 Transcribed MEDLINE Interviews.

Earlier it was noted that more information-giving and question-asking was present in terminal searches. Profiles of percent of searcher interaction by category, for total searcher interaction by interview type (Figure 10) show fairly similar proportional patterns for searcher interaction. Searchers in non-terminal interviews exhibited more proportional activity in the agreement category. Searchers in terminal interviews exhibited more activity in asking for orientation, or facts from the user, but somewhat less in asking for opinions or suggestions. Terminal interviews appear to create more tension or withdrawal situations for searchers than do non-terminal searches.

Users, like searchers, also have similar interaction profiles in the two types of interview (Figure 11). Users in terminal interviews exhibit a somewhat higher proportional activity in the positive (Categories 1 through 3) and negative categories (10 through 12) than do users in non-terminal searches.

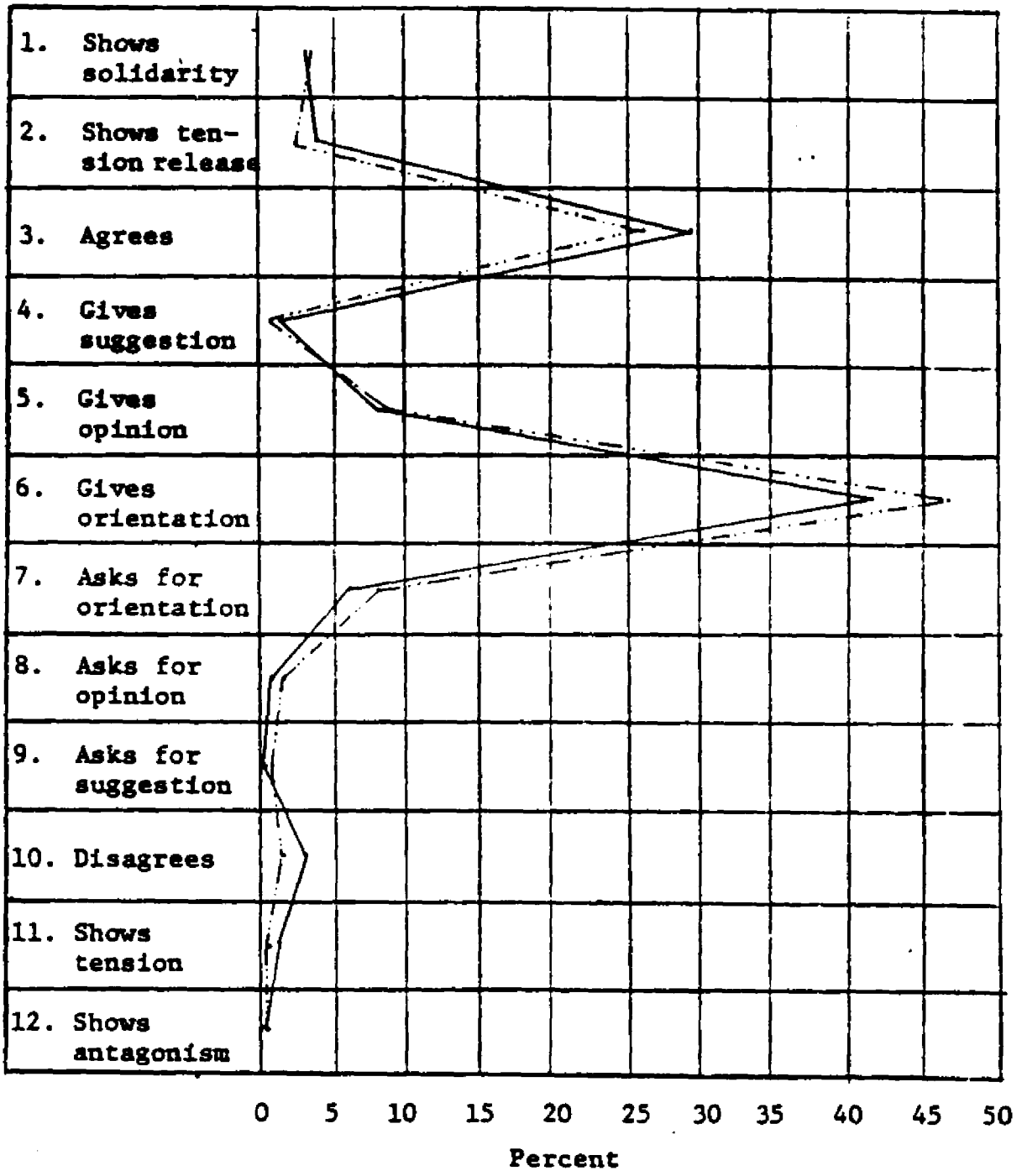
Although reference negotiation has been considered as a process in which the intermediary "interrogates" the user (Taylor, 1968), the profiles outlined here appear to suggest that information-giving is the prime activity for both searchers and users participating in an on-line interview. In a quantitative sense more information-giving by the user and more question-asking by the searcher appear to occur in interviews which are combined with the actual terminal



——— Searcher interaction terminal interviews (N = 2374)

—·—·— Searcher interaction non-terminal interviews (N = 1449)

Figure 10. Searcher interaction by category as percent of total searcher interaction by type of interview.



User interaction terminal interviews (N = 1626)

User interaction non-terminal interviews (N = 1109)

Figure 11. User interaction by category as percent of total user interaction by type of interview.

search. Whether the increased activity is a function of specific events associated with terminal sessions or only the increased amount of time the interviewer and searcher spend together is not readily apparent. However, terminal sessions may provide the atmosphere for more helpful information exchanges simply because they imply an extended interview session. From a subjective perspective it seemed that users in terminal sessions volunteered several important background characteristics about themselves either while waiting for the terminal to respond to some particular input, or as a result of particular citations they viewed in the session. Searchers appeared more likely to question users about seemingly dissonant responses to a particular citation, e.g., asking why a citation was not relevant when it appeared consistent with something the user had previously asked for, or alternatively, asking the user why he liked a particular citation when it did not seem to be within the scope of material previously discussed.

Summary

Tests of the 12 experimental hypotheses advanced in Chapter 1 were considered in this chapter. Moderate relationships were observed for four evaluation areas from the user evaluation form, when each of these areas were considered in conjunction with the overall value designation reported by the user. No significant rank differences in the direction predicted were noted on evaluation measures or information-giving for student and faculty user groups. A significant

relationship between user information-giving and relevance scores was observed. Question-asking by the searcher appeared to enhance this relationship. No significant relationship was noted for user information-giving and the concern for recall score. The last hypothesis was not tested because the primary relationship upon which it depended was found to be non-significant.

Chapter 5 of this report presents an overall summary, conclusions and recommendations for further study.

Chapter 5

Summary and Conclusions

In this study several factors related to user assessments of search results and to events occurring in the search interviews were examined. In a primary sense the study focused on four questions:

1. What relationships exist between the user designated value of the search and other designations associated with the search results and the search interview?
2. Do specified user groups differ in evaluative assessments and in their participation in the search interview?
3. Is information-giving activity by a user in a search interview related to two traditional foci of assessment in search evaluation (relevance and recall)?
4. Is question-asking by the searcher during the interview a contributory factor in any observed relationship between user information-giving and the relevance or recall concern score?

A secondary aspect addressed was the consideration of the interaction profiles developed from a number of interviews coded by Bale's Interaction Process Analysis system.

Sixty-one MEDLINE users meeting study criteria (faculty with a research purpose, students) were sent an evaluation form addressing aspects of the search results and the search

interview. Fifty-four (88%) of the forms were returned within time-limits established for the study.

Eighteen interviews (10 students, 8 faculty) were selected for transcription and subsequent coding by Interaction Process Analysis. Interactive units of primary interest in this coding were the sum of user activity noted in categories 4, 5, and 6, and the sum of searcher activity noted for categories 7, 8, and 9 (information-giving units and question-asking units).

Users generally considered the results of the search to be valuable in meeting the need prompting the search request. Relationships (in decreasing order) were demonstrated between the value designation and the user's satisfaction with the proportion of relevant citations, the recall concern score, the relevance score, and the perception of the searcher in the interview.

Faculty users in the study did not make more stringent assessments in the areas predicted (relevance, concern for recall, or assessment of the search interview), nor did they exhibit significantly more information-giving activity in the search interview. As anticipated no significant differences in value designations and satisfaction with the proportion of relevant citations retrieved was exhibited by faculty and student users studied.

User information-giving activity during the search interview, measured by Interaction Process Analysis, was

related to relevance scores assigned to the searches; a similar relationship between user information-giving activity and the recall concern score was not demonstrated. Question-asking activity by the searcher, measured by Interaction Process Analysis, was related to information-giving behavior by the user. An apparent relationship between searcher question-asking and the relationship of user information-giving to relevance scores was suggested.

Conclusions resulting from this descriptive study emanate from several sources, from data tests; from trends evident in supplementary analysis of the data, but not specifically tested; and from subjective impressions developed by the investigator in the process of the study. The primary limiting factor affecting generalizability to other similar academic MEDLINE search sites, would appear to be the self-selection of users who participated in this study. Such selection is likely to be a continuing factor in the future for similar studies as very proper concerns for the privacy of individual subjects are expressed through the application of increasingly rigid procedures to protect subject rights. For example, users in this study were required to sign a consent form which clearly indicated that, as subjects, they might be asked to participate in an evaluative process in the future. One can speculate that users "in a hurry" were also unlikely to commit themselves to a future investment in time. In contrast, library users in the immediacy, non-immediacy study (Gothberg, 1975) were unaware that any

recording process was being carried out. Users in the Georgia-UCLA study (Carmon, 1975) were asked to give verbal consent to recording of the interview session. Users in Lynch's study (1977) were apprised that the reference session was being recorded by means of a sign on the reference desk and a badge worn by the reference librarian. These users would have had to take the initiative to select themselves out of the study rather than into it.

For users in this study the satisfaction with the proportion of relevant citations retrieved, and expressions of concern in the area of recall for the search, appeared to be more importantly related to the value of the search results than did the actual proportion itself (the relevance score). This suggests that greater elucidation of the individual user's tolerance level for irrelevant citations at the outset of the search process may be important. This aspect is sometimes covered by a broad-narrow question on the request form (e.g., Do you want a broad search retrieving many of the relevant citations, but which may also retrieve many irrelevant citations? or, Do you want a narrow search retrieving primarily relevant citations, with few irrelevant citations, but which may exclude some relevant citations?) A question of this type would also appear to have some bearing on establishing the user's potential concern for recall. This type of question was not included on the request form for this study, so no specific examination of

pre-search expressions of preference in these areas can be made. Specific elucidation of the user's preference during the interview is suggested because, in general, for interviews considered in this study, the user request form, if referred to at all, appeared to serve primarily as a source of information concerning the subject of the search. In many cases the form was not completed by the user until the end of the interview.

By the tests employed for detecting directional differences predicted for faculty in several areas (relevance, concern for recall, perception of the search interview, and information-giving activity in the interview), one must accept the null hypothesis of no difference in user groups (faculty, students) in the absence of evidence supportive of the predictions. However, the trend in the data indicates contrary responses from students in all areas, particularly in the mean rank differences on concern for recall. Some further investigation of student assessments may be appropriate. The results from this study suggest that the presumed greater knowledgeability of faculty (with some justification for the presumption of knowledgeability evidenced in the higher degree levels attained by faculty, the greater likelihood of faculty to read on a regular basis journals related to the search topic, and the greater likelihood that they had published on the topic) is not a factor

leading to lower evaluative assessments or more information-giving.

Interaction profiles developed in this study showed that information-giving activity (as defined by Interaction Process Analysis) constituted the largest proportion of all searcher activity. This observation repeats in a sense the observation by Lynch that "participating librarians [in reference interviews in public libraries] spent a significant proportion of their time giving information to patrons rather than getting information from them" (1977, p. 129). No direct examinations of searcher information-giving and any assessment measures were made in this study. However, it does appear that some relationship exists between user information-giving activity and relevance scores. Question-asking by the searcher has been suggested as a contributory factor in this relationship.

Users in terminal session interviews exhibited more information-giving activity than those in non-terminal interviews. Searchers in terminal interview sessions exhibited more question-asking activity than did searchers in non-terminal sessions. A trend in the data results suggests that users participating in terminal sessions assigned somewhat higher value scores, relevance scores, satisfaction with the proportion of relevant citations scores, and search interview scores than did users in non-terminal sessions. A subjective impression received by the investigator in

listening to the tapes of the interviews is that the user's viewing of actual citations during the interview prompted interactive activity (information-giving, agreements, disagreements, and question-asking) that may have been helpful in the search process. A further examination of the type of interview (terminal, non-terminal) and assessment measures may provide additional information for improving the interactive and diagnostic process.

APPENDIX A

FORMS

OAKLAND UNIVERSITY Rochester, Michigan 48063

KRESGE LIBRARY

A research study supported by the Office of Libraries and Learning Resources, Office of Education is being conducted at the Houston Academy of Medicine - Texas Medical Center Library, the University of Michigan and Wayne State University. This project is directed to an investigation of two areas: 1) the relationship of user characteristics and search purpose with the user judgement of MEDLINE search results and 2) an examination of the MEDLINE search interview for the identification of variables related to user judgment of search results. Primary data collected for this study include user information supplied on a MEDLINE request form; tape recorded interviews for MEDLINE searches; typed copies (all identifying characteristics deleted) of interviews from respondents completing a user evaluation form; and user evaluation forms returned by selected respondents. It is anticipated that the results of this study may provide a basis for clarification of performance objectives for on-line information retrieval systems. As a current user, and a potential future user, of such systems your cooperation in this project is requested.

Your participation involves three areas:

1. Completion of a search request form.
2. Participation in a tape-recorded interview session. Should you agree to recording, you may change your mind at any time during the interview and request that taping cease.
3. Potential selection as a respondent for a twelve-item evaluation form. Items include eleven scale responses and a count of relevant citations in your print-out.

The identities of all participants in this project will be held confidential by the investigator.

Participation in this project is voluntary, non-participation will not affect treatment of your MEDLINE request. Participants may withdraw from the project at any time.

September 6, 1978 Eileen E. Hitchingham

Eileen E. Hitchingham
Project Investigator

I AM WILLING TO PARTICIPATE IN THE PROJECT OUTLINED ABOVE.

Requester _____ Searcher _____

Date _____ Date _____

IN THE FUTURE, IF OTHER QUALIFIED INVESTIGATORS SHOULD REQUEST COPIES OF INTERVIEW TAPES AND/OR COMPLETED REQUEST FORMS, I AGREE TO THEIR RELEASE BY THE PROJECT INVESTIGATOR IF ALL NAME AND INSTITUTION REFERENCES ARE DELETED.

Requester _____ Searcher _____

Date _____ Date _____

k
9/78

Date	___ / ___ / ___ (Mo / Day / Year)
Searcher	_____
Agree to Tape	_____ Yes ____ No
Search Number	_____
Citations	___ On-line ____ Off-line

MEDLINE SEARCH REQUEST FORM

PLEASE PRINT

1. Name _____

2. Telephone Number _____ / _____ / _____
   Area Code / Number / Extension

3. Mailing Address _____

   _____
   City / State / Zip

4. Primary Institutional Affiliation _____

   Department _____

5. Position or Title _____

6. Number of previous on-line search interviews.
   - None _____
   - 1 - 5 _____
   - 6 - 10 _____
   - More than 10 _____

7. The results of this search are primarily for my own use.
   - Yes _____
   - No _____

8. If NO, please indicate Name, Title and Address of person for whom Search is intended:

   Name _____

   Title _____

   Address _____

   _____
   City / State / Zip

9. When search is completed: Send results by mail _____

I will pick up results _____

10. Check the ONE blank which BEST describes the requester of the search. If two designations appear to be applicable (e.g. a Physician who is also a Graduate student) check the status which relates MOST to your request for a search.

EXAMPLE:

A Physician/Graduate student initiating a MEDLINE request PRIMARILY for an actual patient care problem would designate "Physician."

A Physician/Graduate student initiating a MEDLINE request PRIMARILY for graduate-related coursework would indicate "Graduate Student."

_____ 1. Undergraduate student

_____ 2. Medical student

_____ 3. Graduate student (Master's level)

_____ 4. Graduate student (Ph.D. level)

_____ 5. Instructor (includes Clinical & Adjunct appointments)

_____ 6. Ass't. Professor (includes Clinical & Adjunct appointments)

_____ 7. Assoc. Professor (includes Clinical & Adjunct appointments)

_____ 8. Professor (includes Clinical & Adjunct appointments)

_____ 9. Postdoctoral appointment

_____ 10. Intern/Resident

_____ 11. Academic researcher

_____ 12. Physician

_____ 13. Nurse

_____ 14. Dentist

_____ 15. Pharmacist

_____ 16. Other (please specify) _____

11. Check the ONE blank which BEST describes your purpose in requesting a MEDLINE search.

_____ 1. Grant project (development, in progress, completion)
_____ 2. Preparing an article based on research work for publication
_____ 3. Preparing a review article for publication
_____ 4. On-going research which will lead to publication
_____ 5. Dissertation (Doctoral degree)
_____ 6. Thesis (Master's degree)
_____ 7. Term paper
_____ 8. Class project
_____ 9. Background material for a seminar
_____ 10. Background material for a speech or talk
_____ 11. Instruction or teaching
_____ 12. Clinical application (diagnosis, treatment of a patient or client)
_____ 13. Personal bibliography, no immediate application
_____ 14. Other (please state) _____

12. Check ALL degrees completed

_____ 1. High school
_____ 2. Associate's degree
_____ 3. Bachelor's degree
_____ 4. Master's degree
_____ 5. Doctoral degree
_____ 6. M.D.
_____ 7. D.D.S.
_____ 8. Other (please state) _____

13. Please consider the following statements as they apply to the topic of your Search request.

1. I am aware of several recent (last 2 years) publications related to the topic of my Search request.	1.	Yes _____ No _____
2. I read on a regular basis (weekly, monthly) several journals which have articles related to the topic of my request.	2.	Yes _____ No _____
3. I have published an article or articles related to the topic of my Search request.	3.	Yes _____ No _____
4. The topic of this search request is likely to be of continuing interest to me (next one or two years).	4.	Yes _____ No _____

14. Search request: Please give a detailed statement of the subject matter for which the Search is to be conducted. Define any terms which may have special meaning relative to your Search and/or any concepts you wish to exclude.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

15. Relevant citations: If possible list author, title and publication data for any known relevant articles published within the last three years. The citations will be used as a guide for retrieving similar citations related to your needs.

1. Author _____
   Title _____
   Publication Data _____

2. Author _____
   Title _____
   Publication Data _____

3. Author _____
   Title _____
   Publication Data _____

For Use in WAYNE Searches Only.

FOR LIBRARY USE ONLY. DO NOT WRITE IN THIS BOX
Part 2
Library Data

☐ WSU formulated search
Search unit(s) _____
Line time _____
Form time _____

☐ Pre-formulated search
Search unit _____
Line charges _____

☐ Institution run search
Minutes on-line _____

☐ Recurring search
☐ Monthly ☐ Quarterly

Method of payment
☐ Free ☐ Check($_____
☐ IRB ☐ Cash ($_____
☐ To be invoiced

Search formulated by _____
Search run by _____
Date run _____ ☐ NLM ☐
Taskname _____

Total off-line prints:

MEDLINE	_____ @ __ per cit.		TOXLINE	_____ @ __ per cit
BACK72	_____ @ __ per cit.		TOXBACK	_____ @ __ per cit
BACK69	_____ @ __ per cit.		CANCERLINE	_____ @ __ per cit
BACK66	_____ @ __ per cit.		EPILEPSY ABSTS.	_____ @ __ per cit
OTHER (specify)	_____			@ __ per cit

Specify files to be searched:

- ☐ MEDLINE (latest 2-3 years)
- ☐ BACK72 (1972-74)
- ☐ BACK69 (1969-71)
- ☐ BACK66 (1966-68)
- ☐ TOXLINE (1971 to present)
- ☐ TOXBACK (1940-70)
- ☐ CANCERLINE (1963 to present)
- ☐ EPILEPSY ABSTS. (1945 to present)

☐ Other (specify) _____

_____

Search limitations: Please check all boxes and supply information pertinent to the scope of your search.

1. ☐ No restrictions

2. ☐ Human only

   a. ☐ Male ☐ Female

   b. ☐ Infant, newborn (to 1 mo.)
      ☐ Infant (1-23 mos.)
      ☐ Child, preschool (2-5 yrs.)
      ☐ Child (6-12 yrs.)
      ☐ Adolescence (13-18 yrs.)
      ☐ Adult (19-44 yrs.)
      ☐ Middle age (45-64 yrs.)
      ☐ Aged (65 yrs. and older)

3. ☐ Animal only. If only certain animals or animal groups, list below:

   a. _____ d. _____

   b. _____ e. _____

   c. _____ f. _____

4. Language restrictions:

   ☐ Accept all languages

   ☐ English

   ☐ English abstracts

   Other languages accepted (specify):

   _____

   _____

Search Formulation

Pertinent MeSH headings or keywords:

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____
17. _____
18. _____
19. _____
20. _____

Search statements: FOR LIBRARY USE ONLY.

		Pstg.	Off-Line
1.			
2.			
3.			
4.			
5.			
6.			
7.			

Searcher	_____
Search No.	_____

MEDLINE USER RESPONSE FORM

You have been selected from a number of users who participated in a taped interview for a recent MEDLINE search. Your cooperation in completing this form as part of the follow-up analysis of user judgment of search results is sought.

The questions consider four areas:

1. Relevance -- Your judgment of each citation retrieved in the MEDLINE search and its logical connection (relationship) to the question you posed to the system.

2. Recall -- Your judgment of the "completeness" of the search, i.e., are you satisfied that the search retrieved all of the relevant citations in the MEDLINE data base?

3. Value of the Search -- How valuable was the search in meeting the information need which prompted your request?

4. Search Interview -- Your judgment concerning the ability of the searcher to assess and explore all aspects of your information need during the search interview.

Answer ALL questions as best you can and feel free to make any comments you think necessary. Your answers and comments will be kept confidential.

Please complete this form and return it within two weeks to:

Eileen E. Hitchingham
Kresge Library
Oakland University
Rochester Michigan 48063

A stamped envelope is enclosed for your use.

Please turn to questions on the following pages.

rev3

I. RELEVANT

1. Complete this question ONLY if search results (citation print-out) are no longer available to count the number of relevant citations. If the print-out is available GO TO QUESTION 2.

   My estimate of the percent of relevant (related to my search question) citations retrieved in my search is:

0%	_____
1 - 10%	_____
11 - 20%	_____
21 - 30%	_____
31 - 40%	_____
41 - 50%	_____
51 - 60%	_____
61 - 70%	_____
71 - 80%	_____
81 - 90%	_____
91 - 100%	_____

(GO TO QUESTION 3)

2. Your search retrieved _____ citations. Review each citation in your print-out and count the total number of relevant citations. Relevant citations are ALL citations which you consider to be related to your search question, thus, relevant citations include those with which you were already familiar prior to the search as well as new citations related to your question.

   Number of relevant
   Citations in the print-out _____

   (GO TO QUESTION 3)

3. Indicate your degree of satisfaction with the proportion of relevant citations retrieved in the search by marking an X on the appropriate space below.

Unsatisfactory	___	___	___	___	___	___	___	___	___	___	Satisfactory
	1	2	3	4	5	6	7	8	9	10	

(COMPLETE ALL FOLLOWING QUESTIONS)

II. RECALL

Indicate your level of agreement with the following statements by marking an X in the appropriate space.

1. I believe that the search retrieved most of the relevant citations in the MEDLINE data base.

   Disagree ___:___:___:___:___:___:___:___:___:___ Agree
   1 2 3 4 5 6 7 8 9 10

2. In assessing my search results I am concerned because there is no way to judge completeness.

   Disagree ___:___:___:___:___:___:___:___:___:___ Agree
   1 2 3 4 5 6 7 8 9 10

3. I am concerned because the search results omitted relevant citations with which I was familiar prior to the search.

   Disagree ___:___:___:___:___:___:___:___:___:___ Agree
   1 2 3 4 5 6 7 8 9 10

4. The search results included fewer relevant citations than I expected.

   Disagree ___:___:___:___:___:___:___:___:___:___ Agree
   1 2 3 4 5 6 7 8 9 10

III. SEARCH VALUE

1. Consider the purpose for which you requested this search. Indicate your value assessment of the search results (the ability of the search to meet the need prompting your request) by marking an X in the appropriate space.

   No Value ___:___:___:___:___:___:___:___:___:___ Major Value
   1 2 3 4 5 6 7 8 9 10

IV. SEARCH INTERVIEW

Indicate your level of agreement with the following statements by marking an X in the appropriate space.

1. The searcher was knowledgeable concerning the use of the MEDLINE data base for my question.

   Disagree ___:___:___:___:___:___:___:___:___:___ Agree
   1 2 3 4 5 6 7 8 9 10

2. I feel that the searcher understood my request after the interview.

Disagree ___:___:___:___:___:___:___:___:___:___ Agree
1 2 3 4 5 6 7 8 9 10

3. The searcher understood my purpose in initiating the request.

Disagree ___:___:___:___:___:___:___:___:___:___ Agree
1 2 3 4 5 6 7 8 9 10

4. The searcher was thorough in exploring all aspects of my search question.

Disagree ___:___:___:___:___:___:___:___:___:___ Agree
1 2 3 4 5 6 7 8 9 10

5. The searcher suggested terms appropriate to the subject of my request.

Disagree ___:___:___:___:___:___:___:___:___:___ Agree
1 2 3 4 5 6 7 8 9 10

V. COMMENTS

_____

_____

_____

_____

_____

_____

THANK YOU FOR YOUR COOPERATION IN COMPLETING THIS FORM.

Do you wish to receive a report on the final project results?

Yes ___

No ___

rev3

MEDLINE PROJECT PROCEDURES

TEXAS MEDICAL CENTER LIBRARY

Supplies

Materials needed for the project include:

1. Blank cassettes
2. Tally sheets for MEDLINE Searches not included in the project
3. Consent/Request Forms
4. Tape recorder (available at Texas Medical Center Library)
5. Sample Evaluaton Forms
6. Return envelopes for cassettes and Consent/Request forms

Population to be Recorded

For the purpose of this project it is desirable to seek participation from all users who come to the Library and request a MEDLINE Search.

Tally Sheets for Non-Recorded MEDLINE Searches

Since I would like to know the proportion of recorded MEDLINE Searches to all MEDLINE Searches conducted during the participation period, tally sheets are available to note each MEDLINE Search completed for which there is no recording. On the tally sheet please note reason, date, and searcher for any unrecorded MEDLINE Search you may do.

Sequence of Events

1. Consent

When an individual requests a MEDLINE Search, in person, explain that the Library is participating in a MEDLINE evaluation study and you are inviting them to participate in the project. Give the requester a Consent/Request form for her/his consideration. If the requester agrees, obtain her/his signature(s) on the form. (Some requesters may prefer to sign only the first release, others may sign both statements).

Sequence of Events - continued

2. Recording

Begin recording on Side #1 of the cassette immediately after the consent Form is signed. Indicate:

- Date
- Requester Name
- Time
- Search Number (at top of Consent Form)
- Searcher Name

3. Request Form

a) Have user fill out Request Form down to the dotted line on page 5. Additional pages (6-7) were added for the convenience of Wayne State University; they can be cut off if they are not useful for you).

b) Check to see that all information is completed, particularly name, mailing address and phone number.

4. Interview

a) Conduct interview according to usual method

b) If interview is longer than 30 minutes, turn cassette over and continue recording

c) When interview is completed indicate time and stop recording

d) If interview includes working with the requester at the terminal this interaction should be recorded also.

5. After Interview

a) Label cassette with

- date
- search number
- requester name
- searcher name or initials

b) Check to see that Consent Form has requester signature

c) Give the requester a SAMPLE evaluation form. Indicate that she/he may be contacted in the future to respond to a similar questionnaire. The user is NOT to return the sample, it is for her/his information only.

Sequence of Events - continued

5. After Inverview - continued

   d) Sign Consent Form in searcher signature position. If, for some reason, you do not wish to sign even though the requestor has agreed to participate, it would be helpful if you would note this on the signature form so that I will know that it has not just been forgotten.

6. After Search

   a) If on-line citations are printed, note the number of citations in the box in the upper right corner of the request form.

   b) If off-line citations are sent to the user, note the total number of off-line citations sent to the user (e.g. MEDLINE and nay backfiles).

   c) If both on-line and off-line citations result from the search note both of these numbers in the appropriate spot in the box in the upper right of the request form.

   d) Once a week mail all completed cases (the tape plus the Consent/Request form which notes the number of citations given to the user) in the pre-addressed envelopes.

OAKLAND UNIVERSITY Rochester, Michigan 48063

KRESGE LIBRARY

Dear

As you recall, you recently participated in a taped interview for a MEDLINE Search Request. With an Office of Education Grant I am conducting a study of user evaluation of on-line searches; you have been selected as a MEDLINE user important to this study. I am requesting your support in completing the enclosed questionnaire. Since it is anticipated that this study will clarify some factors involved in user satisfaction with an on-line system, I hope you will contribute a few minutes to this project. The confidentiality of your reply will be maintained.

If you have any questions concerning the form, please call me at the number below. I appreciate your time and cooperation, and am looking forward to receiving your completed questionnaire.

Sincerely,

Eileen E. Hitchingham
On-Line Project Director

313/375-0343

k
enc

APPENDIX B

CODING RELIABILITY

SOME CODING CONVENTIONS

Agreed Upon at Coding Session, August 11, 1978*

Category 1 (solidarity, friendly, status raising)

... Hello's, goodbye's
... make yourself comfortable
... that's alright if you want to
... great, fine, excellent
... you don't have to worry about that.

Category 2 (tension release, break-through)

Laughs, jokes, ... now we're getting somewhere..., enthusiasm, satisfaction, Ah!! when a good citation comes up.

Category 3 (agrees)

Sure, um hum, yeah..., right, OK (when agreement not setting new direction) e.g., OK, /6/ let's look at this part of your question now /4/

- "Oh" as understanding, e.g., one speaker has just explained something
- agreement may be expressed in the negative e.g. You don't want to include this category, right? /7/ No /3/ I don't /6/.

Category 4 (gives suggestion)

"Let's look at", "you might want to consider this aspect", "if there are too many, we can..."

Category 5 (opinion, evaluation analysis)

- it seems, ought to, should, perhaps it may, probably, might have, maybe's
- sometimes include qualifiers... pretty well, quite a lot of, it's rather crummy, they're not too good about that
- projection into future -- it looks like you'll get a lot of citations on that
- N.B. in general -- wish is not the same as want e.g. "I wish the System went back earlier than 1966" /5/. "I want articles from the last three years" /6/.

* individual circumstances, conversational flow, context may, at times, indicate a different interpretation.

Category 6 (giving orientation, information, etc.)

- Repetitious, stating what was said in another way
- Now then, OK (not on agree OK but setting direction)
  Alright then (not an agree but setting direction)
- "you know" reiteration.

Category 7 (asks for information, repetition, confirmation, clarification)

- In general a question that could be answered by a yes, no or fact
- a repetition with a question mark
  - U: I want everything on Nematoda /6/
    S: Nematoda? /7/
    U: Nematoda /6/
  Searcher is asking for confirmation
    U: I want everything on Nematoda /6/
    S: Nematoda /6/
  Searcher is repeating, confirming.

Category 8 (asks for opinion, analysis, evaluation, etc)

- do you think there will be many articles on this? /8/
- asking the person to make an inference about a future situation
- generally freeform -- that is, doesn't limit the other person's response.

Category 9 (asks for suggestion, direction,possible ways of action)

- implies an either/or, or suggestion of alternatives upon which action will be taken.

Category 10 (disagrees)

- can be NO or YES, but.

Category 11 (shows tension, asks for help, withdraws from field)

- Statements marked inaudible (i.e. person is speaking or reading to self) withdrawal
- statements with seven dots ..... indicating a long pause searching for answer
- feelings of frustration, disappointment.

Category 12 (shows antagonism)

- Self defensive statements, putting the blame on someone or something else.

Coding Reliability

Several earlier studies which have involved coding of events occurring in the reference or bibliographic search interview have either made no specific coder reliability determinations (Carmon, 1975), or have reported reliability in terms of simple percent of agreement by coders (Gothberg, 1975; Lynch, 1977). Holsti (1969) notes that this method, although often used, has been criticized because it fails to account for the extent of coder agreement arising from chance. For this reason a method noted by Waxler and Mishler (1966) was utilized in the present study. Cohen's "k" is a coefficient of agreement for nominal scales which expresses the proportion of agreement <u>after</u> chance agreement has been removed from consideration (Cohen, 1960).

The formula for computing k is:

$$k = \frac{fo - fc}{N - fc}$$

where k = the coefficient of agreement

fo = the sum of agreements observed for each category

fc = the sum of agreements expected by chance for each category

and N = the total number of observations

An approximation to the standard error of k is given by:

$$\sigma k = \frac{\sqrt{fo\ (1 - fo/N)}}{N - fc}$$

With a large N ($\geq$ 100) the sampling distribution of k approximates normality, so that confidence limits can be expressed as:

95% confidence limits = $k \pm 1.96\ \sigma_k$

In the present study (N = 129) a k of .77 is noted (95% confidence interval = .67 $\longleftrightarrow$ .87). If the more commonly used percent agreement method were applied to the observations, the percent of agreement would equal the upper limit of the interval (.87).

APPENDIX C

USER REQUESTS: ANALYZED INTERVIEWS

Search Request Statements Presented by Users
with Analyzed Search Interviews

I am interested in recent literature on reproduction in nematodes (an animal phyla) and nematomorphs. In particular the areas of spermatogenesis, oogenesis, and accessory glands in these reproductive processes. Most of these studies should be at the ultrastructural level (i.e., electron microscopy).

Cardiac rehabilitation--(exercise training programs) relative to the following diagnoses: myocardial infarction, angina pectoris, coronary insufficiency. This is to include all articles concerning rehabilitative programs which have been established. It is not to be limited by disciplines involved. All foreign language articles are to be excluded.

Blood flow measurement in 1. ovary 2. uterus 3. testis 4. effect of interuterine devices, and

The effect of educational programs on incidence of VD, heart disease and breast cancer. (E.g., is the reported incidence higher or lower--or same--after patients or public's exposure to ed. of the diseases or states?)

I would like to find articles about chlorocruorins for use in my oral examination and for my dissertation.

Clinical pharmacology of propranolol (uses and indications). Back to 1974.

Project summary: To check current literature for enlightening information, both historical and current, relevant to the areas of exercise electrocardiography. This especially entails any changes in the electrocardiogram, either at rest or during stress, both immediately prior to and one month following myocardial infarction. Of particular interest are changes in EKG after the subject has undergone an exercise program.

Effect of increased plasma K^+ on release of catecholamines from adrenal medulla.

The transport of calcium in muscle tissue as it relates to neuromuscular disease.

The subject matter is on the relationship between apgar scores and length of labor.

Milk composition x age.

Respiratory rate vs. respiratory failure.

*Transport mechanisms and transport of newly synthesized proteins into mitochondria.

---

*From tape; user did not write out a statement.

Partial denture design. Effect of occlusal-functional forces on P. dentures. Use of precision attachment in P.D. Effect on abutments in P.D.

*Epithelial neoplasms.

Dihydrofolate reductase: the enzyme, its chemistry, isolation, etc., and genetics. Also inhibitors, in research, not medical uses.

Health behavior of patients--patient compliance, patient attitudes toward health, patient health beliefs.

Beta adrenergic blocking drugs/exercise. Propanolol. pindolol. Metoprolol. Aebutolol.

---

*From tape; user did not write out a statement.

APPENDIX D

EXAMPLE OF CODED TRANSCRIPT

S-1

S: OK, /6/ if you would tell me what you'd like. /8/

U: I want to get a search on beta adrenergic blocking drugs. /6/

S: Uh huh. /3/

U: ...particularly with reference to exercise /6/ and exercise response from the use of the beta adrenergic blocking drugs. /6/

S: OK, /3/ do you want to list some of those drugs for me? /8/ I don't know if they all get lumped together. /5/

U: Um, well particularly propranalol. /6/

S: P-R-O-P-? /7/

U: R-A-N /6/ Propranalol. /6/ Pindolol /6/ P-I-N-D-O-L-O-L /6/

S: Um hmm /3/

U: Metoprolol /6/ M-E-T-O-P-R-O-L-O-L /6/

S: Let me see if those are listed together. /4/

U: They should be listed under beta adrenergic /6/ or beta adrenoreceptor blocking drugs /6/ or maybe under adrenergic blocking drugs. /6/

S: Yeah /3/ OK, /6/ do you want adrenergic beta reception blockaders? /7/

U: Yes /3/ that's it. /6/

S: OK, /6/ so you want these. /9/ Let me show you what's under there /4/ so then you can decide whether you want them all of 'em or some of them. /4/

U: Um hm. /3/

S: OK, /6/ there are those that get listed. /6/

U: Um hmm. /3/

S And you can have all of them /6/ or some of them. /6/

S-2

U: Um hmm. /3/

S: It goes on /6/

U: Yeah /3/ Um ..... /11/ I probably may as well get, um all of them, /5/ because a lot of papers have, like four of them together. /6/

S: Um hm. /3/ OK. /3/

U: So it's hard to tell from a title which ones are left out, /6/ there are four in particular that I want to look at, /6/ but we may as well get them all. /6/ I particularly want to look at the work that's been done in examining exercise and the use of these drugs /6/ because there will be many more publications /6/ than there will be publications with exercise. /6/

S: Right. /3/ Right. /3/ There's usually a lot written on ... /5/ Now /6/ you're just interested in somebody who has ingested this drug and the effect of exercise on its metabolism? /7/ or what? /8/

U: Well the effects of exercise upon blood pressure /6/ and pulse, /6/ exercise tolerance, /6/ anything like that. /6/

S: OK. /3/

U: But they won't have that in the title. /6/

S: Well if I just take any, any, um, incidence of the word exercise appearing with any of these? /7/

U: Yeah. /3/ That should, /5/ that's right, /6/ the title should have, very often, in the majority of incidences, should contain the word exercise. /5/

S: Um hmm /3/ What is the response to exercise? /7/ Is it metabolized differently? /7/ or? /8/

U: Well no, /10/ it affects maximum pulse rate /6/ and the blood pressure response. /6/

S: OK. /3/ How far back do you want to go back on this? /8/

U: Um, well, you have different categories don't you? /8/

S: Right. /3/ The current file is January 76 forward /6/ but the file itself can go back as far as 1966. /6/

S-3

U: Um, what's the differential in price between the ...? /7/

S: OK, /6/ the current file is either a $5.00 /6/ or a $7.00 search /6/ depending on whether or not you want abstracts /6/ if you go all the way back it's $18.00, /6/ or $22.00 if you want abstracts, /6/ and then in between it ranges between those two /6/ so if you want to go back to 1972 /6/ it's $14.00 /6/ or $18.00 depending on the abstracts. /6/

U: Let's go back to '72 /4/ I have a paper with my account number on it. /6/

S: OK, /3/ OK. /3/ And do you want um human and animal both? /9/ Are you going to be using...? /8/

U: Uh, just human. /6/

S: Do you want just English? /7/ or all the languages? /9/

U: Just English. /6/

S: OK. /3/ Do you have a sense of how much stuff is being written on this? /8/

U: Um..... /11/

S: Like 50 articles a year? /9/ or more? /9/ or less? /9/

U: I imagine it would be considerable, /5/ it probably would be 50 articles a year at least. /5/

S: OK. /3/

U: Maybe more than that. /5/ Some of the drugs have been studied much more intensively than others. /6/

S: Um hmm. /3/ OK. /3/ OK, /6/ and your affiliation is with ...? /7/

U: (gives affiliation) /6/

S: Department? /7/

U: Medicine. /6/

S: OK. /6/ And you have an account number there? /7/

S-4

U: This. /6/

S: OK, /6/ what's a phone number I can reach you at? /7/

U: (gives number) /6/ Now I'm going to be away later in the week, /6/ but my secretary can come and get it. /6/

S: OK, /6/ This takes about a week or 10 days /6/ unless you want the first 10 like tomorrow afternoon. /6/

U: That's OK, /3/ I'm going to be away /6/ so next week is fine. /1/

S: OK. /3/ Did you want abstracts with it? /7/

U: No /10/ No. /10/

S: OK. /3/ OK. /3/

U: That would make it too numerous like /6/

S: My name is _____ if you need to get back to me in the next day or so. /6/

U: OK. /3/ Right. /3/

S: If I have a problem you'll be around tomorrow? /7/

U: Yes. /3/ till Thursday. /6/

S: OK, /6/ I'll just get back to you before then. /6/

U: OK. /3/

S: If it doesn't seem quite... /5/ OK, /6/ and thank you for ... /1/

U: OK, /3/ Thanks. /1/

Bibliography

Bales, R. F. Interaction Process Analysis. Cambrdige, Mass.: Addison-Wesley Press, 1950.

Bales, R. F. Interaction process analysis. In D. L. Sills (Ed.), International Encyclopedia of the Social Sciences (Vol. 7). New York: Macmillan, 1968.

Bales, R. F., & Hare, P. A. Diagnostic Use of the Interaction Profile. Journal of Social Psychology, 1965, 67, 239-258.

Benenfeld, A. R., Marcus, R. S., Pensyl, M.E., & Reintjes, J. F. NASIC at MIT, Final report, 1 March 1974 - 28 February 1975 (ESL-FR-587). Cambridge, Mass.: Massachusetts Institute of Technology, February 1975. (ERIC Document Reproduction No. ED 107 226)

Berelson, B. Content analysis. In G. Lindzey (Ed.), Handbook of Social Psychology (Vol. 1). Reading, Mass.: Addison-Wesley, 1954.

Borgatta, E. F. & Crowther, B. A workbook for the study of social interaction processes. Chicago: Rand McNally, 1965.

Cannell, C. F., & Kahn, R. L. Interviewing. In G. Lindzey & E. Aronson (Eds.), The Handbook of Social Psychology (2nd ed., Vol. 2). Reading, Mass.: Addison-Wesley, 1968.

Carmon, J. L. Model the user interface for a multidisciplinary bibliographic information network (GN-41236). Athens, Ga.: University of Georgia, Office of Computing Activities, May 1975. (NTIS No. PB-242 964/56A)

Cohen, J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960, 20, 37-46.

Conover, W. J. Practical nonparametric statistics. New York: Wiley & Sons, 1971.

Cooper, W. S. On selecting a method of retrieval effectiveness. Part I. The subjective philosophy of evaluation. Journal of the American Society for Information Science, 1973, 24, 87-100.

Daniels, L. A matter of form. Online, 1978, 2(4), 31-39.

Fosdick, H. An SDC-based on-line search service, a patron evaluation survey and implementation. Special Libraries, 1977, 68, 305-312.

Francillon, M. Information retrieval: A view from the reference desk. Journal of Documentation, 1959, 15, 187-198.

Gothberg, H. User satisfaction with a librarian's immediate and nonimmediate verbal-nonverbal communication (Doctoral dissertation, University of Denver, 1974). Dissertation Abstracts International, 1975, 35, 6839A. (University Microfilms No. 75-05025)

Gothberg, H. Immediacy: A study of communication effect on the reference process. Journal of Academic Librarianship, 1976, 2, 126-129.

Hawkins, D. T. Online information retrieval bibliography. Online Review, 1978, 2, 63-106.

Heyns, R. W. & Lippit, R. Systematic observational techniques. In G. Lindzey (Ed.), Handbook of Social Psychology (Vol. 1). Reading, Mass.: Addison-Wesley, 1954.

Hitchingham, E. E. MEDLINE use in a university without a school of medicine. Special Libraries, 1976, 67, 188-194.

Holsti, O. R. Content analysis. In G. Lindzey & E. Aronson (Eds.), The Handbook of Social Psychology (2nd ed., Vol. 2). Reading, Mass.: Addison-Wesley, 1968.

Holsti, O. R. Content analysis for the social sciences and humanities. Reading, Mass.: Addison-Wesley, 1969.

Hoover, R. E. Patron appraisal of computer-aided on line bibliographic retrieval services. Journal of Library Automation, 1976, 9, 335-350.

Jahoda, G. Instruction in negotiating the reference query. Tallahasee, Fla.: School of Library Science, Florida State University, August 1975. (ERIC Document Reproduction Service No. ED 111 421)

Jahoda, G. The process of answering reference questions. A test of a descriptive model. Tallahasee, Fla.: School of Library Science, Florida State University, January, 1977. (ERIC Document Reproduction Service No. ED 136 769)

Jahoda, G., Bayer, A., & Needham, W. L. A comparison of on-line bibliographic searches in one academic and one industrial organization. RQ, 1978, 18, 42-49.

Kobelski, P. & Trumbore, J. Student use of online bibliographic services. Journal of Academic Librarianship, 1978, 4, 14-18.

Lancaster, F. W. Evaluation of the MEDLARS demand search service (National Library of Medicine, Health Education and Welfare Dept.). Washington, D.C.: U. S. Government Printing Office, 1968. (S.O.D. No. FS 2.202: M46/2)

Lancaster, F. W. Evaluation of on-line searching in MEDLARS (AIM-TWX) by biomedical practicioners. Occasional Papers of the University of Illinois Graduate School of Library Science, 1972 (No. 101).

Lawrence, B., Weil, B. H., & Graham, M. H. Making on-line search available in an industrial research environment. Journal of the American Society for Information Science, 1974, 25, 364-369.

Librarians on MARS: session probes new machine assisted reference services. American Libraries, 1974, 7, 504.

Link, A. D. A comparison of retrieval efficiency from the Educational Resources Information Center (ERIC) automated information retrieval system using computer-assisted-instruction training and search negotiations as requester-to-system interface methods. Doctoral dissertation, New Mexico State University, 1972. (ERIC Document Reproduction No. ED 067 215)

Lynch, M. J. Reference interviews in public libraries (Doctoral dissertation, Rutgers University, 1977). Dissertation Abstracts International, 1977, 38, 530A. (University Microfilms No. 77-17, 547)

Lynch, M. J. Reference interviews in public libraries. Library Quarterly, 1978, 48, 119-142.

McCarn, D. & Leiter, J. On-line services in medicine and beyond. Science, 1973, 181, 318-324.

McCarthy, S. E., Feng, C. H., & Maccabee, S. S. Evaluation of MEDLINE service by user survey. Bulletin of the Medical Library Association, 1974, 62, 367-373.

Mick, C. K. Investigation of the public library as a linking agent to major scientific, educational, social and environmental data bases: Final evaluation report. Stanford, Calif.: Applied Communication Research, September 1977. (ERIC Document Reproduction No. ED 148 347)

Moll, W. Observations on the AIM-TWX service at the University of Virginia Medical Library. Bulletin of the Medical Library Association, 1972, 60, 571-574.

Nunnally, J. C. Psychometric theory. New York: McGraw-Hill, 1967.

Nie, N., Hull, C., Jenkins, J., Steinbrenner, K., & Bent, D. SPSS Statistical Package for the Social Sciences (2nd ed.). New York: McGraw Hill, 1975.

Nyren, K. The online revolution in libraries. Library Journal, February 15, 1978, pp. 439-441.

Pearse, W. H., Flora, R. E., Freeman, R., & Peeples, E. H. Medical school faculty productivity. Journal of Medical Education, 1976, 51, 201-202.

Rees, A. M., & Schultz, D. G. A field experimental approach to the study of relevance assessments in relation to document searching (2 vols.). Cleveland, Ohio: Case Western Reserve University, Center for Documentation and Communication Research, October 1967. (NTIS No. PB-176-079, PB-176-080)

Saracevic, T. The concept of "relevance" in information science: A historical review. In T. Saracevic (Ed.), Introduction to Information Science. New York: Bowker, 1970. (a)

Saracevic, T. Selected results from an inquiry into testing of information retrieval systems. In T. Saracevic (Ed.), Introduction to Information Science. New York: Bowker, 1970. (b)

Saracevic, T. Relevance: A review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science, 1975, 26, 321-343.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Somerville, A. N. The place of the reference interview in computer searching: The academic setting. Online, 1977, 1, 14-23.

Summit, R. K. & Firschein, O. Two year interim report. Investigation of the public library as a linking agent to major scientific, educational, social and environmental data bases. Palo Alto, Ca.: Lockheed, September 1976. (ERIC Document Reproduction No. ED 131 857)

Tagliacozzo, R. The consumers of new information technology: A survey of the utilization of MEDLINE. Journal of the American Society for Information Science, 1975, 26, 294-304.

Tagliacozzo, R. Estimating the satisfaction of information users. Bulletin of the Medical Library Association, 1977, 65, 243-249.

Taylor, R. S. Question-negotiation and information seeking in libraries. College and Research Libraries, 1968, 29, 178-194.

Tessier, J., Atherton, P., & Crouch, W. New measures of user satisfaction with computer-based literature searches. Special Libraries, 1977, 68, 383-389.

Wanger, J., Cuadra, C. A., & Fishburn, M. Impact of on-line retrieval services: A survey of users 1974-75. Santa Monica, Ca.: System Development Corporation, 1976.

Waxler, N. E. & Mishler, E. G. Scoring and reliability problems in Interaction Process Analysis: A methodological note. Sociometry, 1966, 29, 28-40.

Williams, M. 1977 data-base and on-line statistics. Bulletin of the American Society for Information Science, 1977 4(2), 21-23.

Reference Notes

1. Hitchingham, E. E., User evaluation of on-line searching: Categorization of elements in user evaluation forms. A compilation of materials prepared for a poster session at the ASIS Mid-Year Meeting, Rice University, Houston, Texas, May 21-24, 1978. (Available from the author at Kresge Library, Oakland University, Rochester, Michigan 48063.)

2. Jestes, E. C. CAIN On-line project report for the period October-December 1973. Unpublished manuscript, University Library, University of California-Davis, June 1974.

Autobiographical Statement

EILEEN ELIZABETH HITCHINGHAM

BORN: June 29, 1943
Baltimore, Maryland

EDUCATION:

1965 Chestnut Hill College
Philadelphia, Pennsylvania
B.S. in Chemistry

1966 Western Michigan University
Kalamazoo, Michigan
M.S. in Library Science

1979 Wayne State University
Detroit, Michigan
Ph.D. in Education (Evaluation and Research)

EMPLOYMENT:

1967 to 1968 Harvard University
Boston, Massachusetts
MEDLARS Analyst

1968 to present Oakland University
Rochester, Michigan
Librarian

PUBLICATIONS:

Hitchingham, E. E. On-line data bases I. Use of commercially vended on-line data bases by academic libraries. A report for the Council on Library Resources Fellowship Program. Rochester, Michigan: Oakland University, Kresge Library, 1975. (ERIC No. ED 138 260)

_____. On-line data bases II. Search performance evaluation and on-line data bases. A report for the Council on Library Resources Fellowship Program. Rochester, Michigan: Oakland University, Kresge Library, 1975. (ERIC No. ED 139-388)

_____. MEDLINE use in a university without a school of medicine. Special Libraries, 1976, 67, 188-194.

_____. Selecting measures applicable to evaluation of on-line literature searching. Drexel Library Quarterly, 1977, 13, 52-67.