

Accounting for users' inflated assessments of on-line catalogue search performance and usefulness: an experimental study

[Charles R. Hildreth](#)

Palmer School of Library and Information Science
Long Island University
720 Northern Blvd.
Brookville, NY, USA 11548-1300

Abstract

User-oriented approaches to information retrieval (IR) system performance evaluation assign a major role to user satisfaction with search results and overall system performance. Expressed satisfaction with search results is often used as a measure of utility. Many research studies indicate that users of on-line library catalogs (OPACs) and other IR systems often express satisfaction with poor search results. This phenomenon of "false positives," inflated assessments of search results and system performance, has not been adequately explained. Non-performance factors such as interface style and ease of use may have an affect on a searcher's satisfaction with search results. The research described in this report investigates this phenomenon. This paper presents the findings of an experimental study which compared users' search performance and assessments of ease of use, system usefulness, and satisfaction with search results after use of a Web OPAC or its conventional counterpart. The primary questions addressed by this research center on the influence of two experimental factors, OPAC search interface style and search task level of difficulty, on the dependent variables: actual search performance, perceptions of ease of use and system usefulness, and assessments of satisfaction with search results. The study also investigated associations between perceived ease of use, system usefulness, and satisfaction with search results. Lastly, the study looked for associations between the dependent variables and personal characteristics. No association was found between satisfaction with search results and actual search performance. Web OPAC searchers outperformed Text OPAC searchers, but search task level of difficulty is a major determinant of search success. A strong positive correlation was found between perceptions of system ease of use and assessments of search results.

Introduction

Research has shown that users of on-line library computer catalogs (OPACs) and other computer-based information retrieval systems often express satisfaction with their search results and the overall performance of the retrieval system even when the results, upon analysis, are shown to be poor. This has been called the phenomenon of "false positives" in user assessments of search success ([Applegate](#), 1993, 525). A fundamental assumption of the user-oriented approach to information retrieval (IR) system performance testing and evaluation is that end users of a system, those individuals who bring information needs and questions to the system, are the best judges of the quality of search results and the performance of the retrieval system. If the phenomenon of false positives, that is, unwarranted user satisfaction with search results, is as common as research and experience indicates, this assumption must be reconsidered. Search results quality and system performance are often interlinked in IR system evaluation studies. Perhaps it is time to reexamine the presumed close association of these two variables. If users commonly misjudge the quality of their search results and express satisfaction with those results, is it not time to

question the use of this criterion and performance measure in evaluation studies? Users may be wrong in their initial assessments of search results, and bad IR systems may produce acceptable results, accidentally or not. At the very least, researchers should continue to examine this disparity between users' perceptions and the realities of system performance with the aim of discovering possible explanations for this problematic phenomenon.

The research described in this paper investigates the phenomenon of false positives in the context of on-line library catalog (OPAC) use. Unlike their conventional text-based command and menu-driven predecessors, Web-based, graphical, hypertext OPACs hold the promise of easy search interaction and navigation among related subjects and related items. Facilitated by easy to navigate linkages and searching that does not require prior knowledge of subject headings or class categories and codes, it might be assumed that searchers would rate these Web OPACs as easier to use, more useful and helpful in the search process, and superior in overall performance. This and related assumptions are examined in this study. The existence and network accessibility of two versions of many OPACs, a Web version and a conventional text version, provided the opportunity to proceed with this investigation.

The study described here featured an experiment that was designed to compare users' search performance and assessments of ease of use, system usefulness, and search success after use of either a Web-based hypertext OPAC or its conventional, text-based counterpart. The study took place in a university setting. The research questions addressed in the study include the following:

1. Why do OPAC users often overrate the success of their searches and the system's performance? Why are users often satisfied with poor search results? How may we account for "false positives" in user assessments of their search results?
2. What interface and affective factors influence users' assessments of system ease of use, system usefulness in assisting with the search and retrieval process, and satisfaction with search results? For example, does perceived ease of use affect assessments of system usefulness and search results?
3. What associations exist between search performance and factors such as perceived ease of use, system usefulness, and satisfaction with search results?
4. What associations may be discovered between actual search performance or perceived ease of use and personal characteristics such as gender, educational level, major area of study, previous OPAC or Web experience?
5. What factors are better predictors of search success, user interface factors, or search task level of difficulty?
6. Does search task level of difficulty influence assessments of search success, perceived ease of use, or ratings of system usefulness?
7. Will WebOPACs be rated as superior in ease of use and search performance independently of search task level of difficulty?
8. Who will benefit most from these new network-accessible, hypertext, GUI Web OPACs?

Satisfaction and Success in on-line Retrieval: The Research Context

The Cranfield indexing experiments in the 1960s are often cited as the beginning of the modern era of computer-based retrieval system evaluation ([Cleverdon, Mills and Keen, 1966](#)). In the Cranfield studies, retrieval experiments were conducted on a variety of test databases in a controlled, laboratory-like setting. In the second series of experiments, known as Cranfield II, alternative indexing languages constituted the performance variable under investigation. The aim of the research was to find ways to improve the relative retrieval effectiveness of IR systems through better indexing languages and methods ([Cleverdon, 1970](#)). The components of the Cranfield experiments were: a small test collection of documents, a set of test queries, and a set of relevance judgments, that is a set of documents judged to be relevant to each query. Human searchers, their interaction with the system, their interpretation of the query, and their process-formed relevance judgments were not factors included in these experiments. For purposes of performance comparisons, it was necessary to select quantitative measures of relevant documents output by the system under various controlled conditions. The measures used in the Cranfield II experiments are recall and precision, derivatives of the concept of relevance. Recall is a measure of the proportion of relevant documents in the database actually retrieved in response to a given query; precision is a measure of the proportion of retrieved documents that are relevant. These performance measures are elegant: they are precise and simple to understand. When recall and precision are computed for different sets of search results, comparisons of the factors or systems that produced the output sets is straightforward and accurate.

The Cranfield experiments soon became the exemplar for evaluation of information retrieval system effectiveness

([Harter and Hert](#), 1997, 8). Because of its reliance on careful experimental conditions and quantitative performance measures, the Cranfield experience is rightly respected for its attempt to place IR system evaluation on a sound scientific basis. Theoretically, the Cranfield model relies almost entirely on the attractive, but troublesome concept of relevance. Furthermore, two key assumptions underlie the Cranfield model: users desire to retrieve documents relevant to their search queries and don't want to see documents not relevant to their queries, and document relevance to a query is an objectively discernible property of the document. Neither of these two assumptions has stood the test of time, experience and astute analysis.

What, then, explains the continued application in IR system evaluation studies of the Cranfield model, however modified by user feedback mechanisms? As Harter and Hert note, "the Cranfield evaluation model has served as the basic experimental model for the majority of IR research conducted during the past 30 years." ([Harter and Hert](#), 1997, 13). If one accepts its key assumptions, in tests for retrieval effectiveness only system properties need to be controlled, manipulated, or optimized to affect system performance as measured by recall or precision. The measurements are made on sets of retrieved documents judged to be relevant prior to the search and retrieval process. Users don't intervene and user-based factors can be ignored. After all, as the Donna Harmon of the Text Retrieval Evaluation Conference (TREC) reports, "the human element in the interactive track compounds the problem immensely" ([Harmon](#), 1996, 22). The Cranfield model is a classic example of the *system-oriented* approach to IR system performance and effectiveness evaluation. Perhaps the model is best understood as a "child" of its time. This was a time of batch processing computer systems. Interactive, end-user, public access computer systems were not generally available in the 1960s. Hypertext systems as computer-based information systems had not yet been invented.

The venerable Cranfield evaluation model and measurement techniques continue to serve as the approach of choice in many IR evaluation studies. The most notable among these is the TREC (Text Retrieval Evaluation Conference) series of retrieval evaluation studies ([Harmon](#), 1995, 1996). Furthermore, the relevance-based measures of recall and precision are the system performance measures most used today. They are used in evaluation studies of conventional on-line database search systems, CD-ROM database search systems, OPACs, and Web search engines.

Almost from the beginning of its acceptance as the evaluation paradigm, the Cranfield model had its critics. In 1973 [Cooper](#) issued his challenge to its fundamental premise, the premise which holds that recall and precision measures are valid measures of IR system performance. There is an assumption within this assumption, namely, that relevance is a context-, task-independent, objective property of documents. In "On Selecting a Measure of Retrieval Effectiveness," Cooper argues that any valid measure of IR system performance must be derived from the primary objective of any such system, namely, to satisfy the specific information needs of a particular user at a given time ([Cooper](#), 1973). Thus, a measure of utility to the user is called for. Based on the alternative assumption that the end-user of the information is the best judge of the utility of retrieved documents, Cooper concludes that *satisfaction* with the search output of a system, subjectively assessed by the user, is a superior measure of system performance.

Cooper's published papers provided the rationale, if not the theoretical underpinnings, for the *user-oriented* approach to IR system evaluation. According to this view, systems should be evaluated on the basis of how well they provide the information needed by a user, and the best judge of the system's performance in doing this is the person who is going to use the information. Subjective user satisfaction with search results is deemed to be the best measure of this utility criterion.

In a well-known critical response to Cooper's papers, Soergel rejects Cooper's proposition that a user's satisfaction with search results is a valid measure of retrieval effectiveness ([Soergel](#), 1976). Soergel agrees that utility should be the central objective and focus of IR system evaluation. Soergel argues that overall improvement in the task performance of the user or resolution of the user's problem is a more appropriate measure of utility than subjective satisfaction with initial search results. Soergel is one of the first researchers to point out that users may be satisfied with less than optimal search results, especially when that assessment is made only at the first moment search results are delivered by the system. More than satisfying the user at this point, "What is needed instead is an attempt to make the user successful." ([Soergel](#), 1976, 257)

Soergel's criticisms of satisfaction with search output as a measure of retrieval performance did little to slow or hinder the adoption and embrace of this measure by researchers committed to a user-centered model of system evaluation. Immediately, efforts were made to rescue the concept from Soergel's assault. Recognizing that satisfaction is a subjective state of mind, "inside the user's head," Tessier et al, argue that "Satisfaction from the user's point of view is important because it can be used by the system manager to determine how well the service

performs." ([Tessier, Crouch and Atherton](#), 1977, 384). In expanding the satisfaction construct, the authors identify four aspects of satisfaction: the user's subjective experience, the library as a whole, the reference service, and the interaction with library personnel. They recognize that this shifts the burden on researchers to devise reliable instruments to measure this multidimensional variable.

Retrieval system performance evaluation came to be recognized as a complex, multidimensional problem, a challenge simply too large and complex for the Cranfield model. Both intermediary and end-user systems were increasing in kind and number. New measures and methods of measurement of a variety of performance factors were needed. These factors include system factors, user factors, task factors, and environmental factors, all factors that relate to overall IR system effectiveness. [Harter and Hert](#) (1997, 34) highlight key evaluation studies that use multiple methods to address multiple aspects of system performance and effectiveness. [Delone and McLean](#) (1992), as one example, identify six dimensions of system effectiveness, each requiring consideration in evaluations of system effectiveness: system quality, information quality, use, user satisfaction, individual impact (Soergel's favorite?), and organizational impact.

Dervin and Nilan articulated the distinction between system-oriented and user-centered evaluation paradigms. They also pleaded with IR system researchers to close the "research gap" (the gap between system factors and users' needs) and provide "guidance for the orientation." ([Dervin and McLean](#), 1986, 8). [Bates](#) (1990) reminded us that identifying the line of demarcation between system factors and user factors is no easy task. Just as the Cranfield model reflected the technology and retrieval environment of the 1960s, the adoption by researchers of the user-centered model and new methods of evaluation consistent with that model reflected the IR technology of the 1980s. This was the decade of the "end user." on-line catalogs (OPACs) and CD-ROM databases were widely and quickly adopted and became commonplace in libraries of all types by the end of the decade. In this environment, the user is both the primary searcher of the system and the user of the information it provides. Who better to target by researchers for assistance with system evaluation? The user became the center of attention in evaluation studies. This caused a few classic Cranfielders to grouse and grumble about the inclusion of this "messy" ingredient in the evaluation mix.

The Cranfield model has not lost its appeal, but researchers now typically employ two or more measurement and evaluation methods in studies of interactive retrieval systems. These include transaction log analysis, questionnaires, interviews, video-based observation, and classic recall and precision analysis. The OPAC research literature is rich with examples of studies which use multiple evaluation methods ([Beaulieu and Borgman](#), 1996). In addition to quality measures of output, several other factors have become the focus of evaluation efforts. Among these are ease of use, system browsability, system efficiency, and satisfaction with the system and the search experience as a whole. Su (1991, 1996) has identified 20 measures of retrieval system performance. Her research supports Soergel's position that utility, defined as the proven value to the user down the line, is a better measure of IR system performance and effectiveness than either recall and precision.

Harter and Hert define two general classes of user-oriented measures used in IR evaluation: measures based on users' perceptions and attitudes, and measures that focus on actual user-system interaction ([Harter and Hert](#), 1997, 36). They review the literature on satisfaction in IR system evaluation, a central focus of the study reported in this paper. In the first class of user-oriented measures are affective measures such as usefulness, ease of use, and satisfaction. These measures focus not so much on system factors, apparent or not, but on users' perceptions and assessments of search results and the interaction experienced in using an IR system to achieve these results. The rationale for using these measures is clearly expressed by [Belkin and Vickery](#) (1986, 192): any "information system ought to be evaluated on the basis of how useful it is to its users." This rationale places a key question into clear focus: if utility is the critical concept to be used in evaluations and comparisons of IR systems, is the system user, in this age of end-user searching, the best judge, in all cases, of the quality of search results and overall system performance? The evaluation studies that rely on measures such as users' perceptions of ease of use and subjective satisfaction with search results do not provide a clear and consistent answer to this question.

Satisfaction, as [Harter and Hert](#) report, "has been the most widely used evaluation concept of this kind." (1997, 37). The authors review both the LIS and MIS literature on the use of the satisfaction construct in information system research and evaluation. [Gluck](#) (1996) provides a complimentary review of the major research on user satisfaction that has appeared in the LIS and MIS literature. [Tonta](#) (1992) reviews the research on search failures, including those studies that use user satisfaction as a measure of search failure or success.

[Belkin and Vickery](#) warned, like [Tessier et al](#) before them, of the many problems associated with the satisfaction

construct. These problems include, how to define satisfaction, and how to measure it. To these must be added questions about the reliability of the satisfaction construct as a measure in any particular study, and its lack of independence from other influential, potentially biasing factors in the retrieval environment mix. When used as a performance measure in IR system evaluation, it may be too easily affected by non-performance factors that confound the results. This concern is especially critical when the dependent performance variable being measured is quality of search results, or assessments by the user of "search success."

The few evaluation studies that use satisfaction as a measure of search success and system retrieval performance have produced inconclusive and conflicting results. This situation may be explained in part by the absence of agreement on a uniform definition of the satisfaction concept, or the use of different measurement instruments. [Hiltz and Johnson](#) (1989) emphasize that the *object* of satisfaction (e.g., search results, ease of use, etc.) must be identified and clearly defined from the start. They urge caution about the dangers of a spillover effect where responses to factors such as ease of use or new technology may pollute a user's assessments of actual search results.

In their study of CD-ROM database searchers, [Steffey and Meyer](#) discovered determinants of satisfaction not associated with actual search success. These potentially confounding variables include experience with computers and fascination with the new database search technology. "Patrons were so pleased with the electronic periodical indexes, that it did not matter how satisfied they were with the number of citations they had retrieved, or with *the value of those citations*." (Steffey and Meyer, 1989, 43, emphasis mine). In her research, [Sandore](#) (1990) focused directly on user's satisfaction with search results. Users were given time after their interactive search session was finished to review their search results before indicating their level of satisfaction with the results. She reports finding a low association between precision and satisfaction. Users were often satisfied with low-precision search result sets, even in cases where their expectations and goal were to achieve high-precision results.

[Ankeny's](#) (1991) study of end-user on-line retrieval services discovered an apparent disparity between user satisfaction with the services and actual search success. The success rate of users in this study was quite low, but these users reported high levels of satisfaction with the on-line search services. Ankeny warns us that future evaluation research must devise and use measurement instruments that are able to make distinctions between competing satisfaction variables.

In their study of OPACs at four different libraries, Crawford and his associates found a positive correlation between overall search satisfaction and search success. Search success was defined as "finding what the user wants." ([Crawford, Arnold, Connolly, and Shylaja](#), 1992, 82) These researchers applied a multidimensional definition to overall search satisfaction, which included user-system interface factors. In retrospect, the authors pose this critical question: "Since the measures of satisfaction and success are so similar and are highly correlated, are studies such as this one actually measuring different variables when they study satisfaction and success?" (85).

[Tonta](#) (1992) reviews the literature on search failures, including those studies that use satisfaction as a measure of search failure. The findings present a mixed picture regarding the reliability of this measure. Tonta urges caution in our interpretations of success and satisfaction study findings. They may differ in the variable measured, and a particular study may fail to account for the influence of non-performance factors on assessments of satisfaction with search results.

In a comparison of CD-ROM database searching by professional librarians and university faculty and graduate students, Lancaster and his associates used recall and precision to examine the search results of the two groups. They discovered that student and faculty searchers found only about one-third of the really important items. The authors are disturbed with the discovery that CD-ROM searchers are usually satisfied with less than optimal search results. The authors add: "Many express satisfaction even when they achieve very poor results." ([Lancaster, Elzy, Zeter, Metzler, and Low](#), 1994, 382). They suggest that users are overly impressed with new electronic retrieval technologies, and that this may account for inflated levels of satisfaction with actual search results. Tinanoff studied CD-ROM searchers in a public library and came to similar conclusions: "The users of CD-ROM database products seem to be satisfied with the products and the technology, but perhaps too easily satisfied." ([Tinanoff](#), 1996, 4)

Gluck examines the relationship between relevance and user satisfaction in IR systems. He claims to have "unconfounded" these two measures of system performance at the retrieved item level. ([Gluck](#), 1996, 89). The study's findings indicate a positive correlation exists between relevance of retrieved items and user satisfaction with those items. Su also found that users' satisfaction with "search results as a whole" correlated with recall, the proportion of relevant items retrieved: "They tended to give low ratings when their searches failed to retrieve

relevant references." ([Su](#), 1996, 236).

Very different findings relating to the relationship between user satisfaction and precision and recall were reported by [Saracevic and Kantor](#) (1988) after their extensive study of on-line database searchers. "Satisfaction with results," one of five utility measures used in the study, measured on a 5-point Likert scale, correlated with precision but not recall. Similar correlations were found with the other utility measures used in the study (for example, whether the retrieved information contributed to the resolution of the user's research problem). The authors conclude that the utility of results will likely be associated with high precision search results, and that recall will be a less significant factor in user assessments.

What accounts for these different and sometimes conflicting research findings? Two explanations may be proposed: lack of agreement in a definition of "user satisfaction" as an evaluation construct, even when it is expressly tied to search results as output of an IR system, and the wide variety of measurement instruments and methods employed in IR system evaluation studies. Today, there is general agreement among researchers that user satisfaction as a utility measure is a many-faceted, multidimensional variable. They also agree that when satisfaction is scrutinized in a particular study, a variety of other utility factors and non-performance factors may exercise an influence on satisfaction.

In retrospect, one finding about users and their search experiences has been reported too often to be ignored or treated lightly. Users of IR systems frequently express satisfaction with poor search results and, perhaps, poor system performance. The evidence requires us to question the validity of these user assessment variables, for example, satisfaction with search results and perceived ease of use, as measures and predictors of actual search success and system performance.

Why do users express satisfaction with poor search results? Applegate tackles this problem she calls the phenomenon of "false positives." "A false positive occurs when a consumer is satisfied with an inferior product." ([Applegate](#), 1993, 525). In the IR environment, a false positive occurs when a user judges a search result to be satisfactory when in fact it is not. Applegate describes three models of user satisfaction, the "material satisfaction model," the "emotional satisfaction – simple path model," and the "emotional satisfaction – multiple path model." The first measures actual product quality. The other two measure subjective impressions and assessments, either along one or several dimensions. Recall and precision are appropriate measures for the material satisfaction model. Happiness or emotional satisfaction with search results and related factors such as search setting and task expectations are the primary measurement constructs of the emotional satisfaction – multiple path model. Applegate believes this model explains the phenomenon of false positives in IR system use, namely, users who are satisfied with bad searches. She suggests that "emotional satisfaction may be determined by something other than material satisfaction." ([Applegate](#), 1993, 526). Furthermore, emotional satisfaction may be a partial indicator of material satisfaction. The interaction between these satisfaction variables is not well understood. Applegate urges researchers to carefully distinguish types of satisfaction and different objects of satisfaction in future studies.

Applegate's paper on false positives in IR provided much of the motivation for undertaking the study described in this paper. The real world environment of the Internet and the Web provided the opportunity to design a comparative study of a Web OPAC and the same OPAC in its pre-Web, conventional incarnation. A major objective of this study was to examine possible explanations for false positives in OPAC use.

Recent OPAC Research

Many accounts of OPAC research have been published in the last 15-20 years. A recent "Special Topic" issue of the Journal of the American Society for Information Science (JASIS) covered state-of-the-art OPAC research ([Beaulieu and Borgman](#), 1996). Citations to many of the best research studies can be found in the articles published in this special issue of JASIS. Hildreth reviews much of this research in an essay on old and new design models for on-line catalogs. ([Hildreth](#), 1995a). An analytical review of recent OPAC research is provided by [Large and Beheshti](#) (1997). The authors focus on the various methodologies employed in OPAC studies, and summarize research-based recommendations under three headings: database record enhancement, search capabilities, and interface design.

Little has been published to date on the performance of Web-based OPACs or user satisfaction with these GUI, hyperlink-capable on-line catalogs. A search for research on Web-based OPACs turned up only a few publications. Apparently, no experimental studies involving Web OPACs have been conducted. [Hildreth](#) (1995b) looked at the new graphical user interfaces (GUIs) that are being applied to older, conventional, second generation OPACs. He

warned that users may be too easily impressed with these systems, systems that deliver the same old level of poor results. In her insightful 1996 article, [Borgman](#) asks, "Why are on-line catalogs still so hard to use?" (Borgman, 1996). Perhaps we should be asking now, "Why do easy OPACs still produce such poor results?"

The presentation of bibliographic information in Web OPACs has been the focus of several recent studies. Cherry reports on her comparative study of bibliographic displays in 12 conventional OPACs and 10 Web OPACs. She developed an index of desirable display features to permit scoring of the two sets of OPACs. When assessed against this checklist of features, Web OPACs scored only slightly higher than pre-Web OPACs (60 percent to 58 percent). ([Cherry](#), 1998, 124). [Ayres, Nielsen, and Ridley](#) (1999) describe "BOPAC2," a research project funded by the British Library Research and Innovation Centre that was designed to test and evaluate a Web front end. This front end facilitates uniform access to a number of different library catalogs via the Z39.50 search and retrieval protocol. The BOPAC2 (the "B" is for Bradford University) research focuses on the management at the user interface of very large and complex retrieval sets. Research is still in progress and updates are provided at this Web site:

<http://www.comp.brad.ac.uk/research/database/bopac2.html> (accessed 3/2/00). Carlyle and Timmons (1999) have recently completed a comparative study of the composition of default displays of bibliographic records in 100 Web OPACs. A report of this study can be found on the Web at:

<http://www.ischool.washington.edu/research/projects.cfm>.

In "Web-based OPACs: Between Tradition and Innovation," [Ortiz-Repiso and Moscoso](#) (1999) report on their analytical study of Web OPACs. The goal of their research is to investigate OPACs available on the Web in order to ascertain how successful these new OPACs are at solving the problems associated with first and second generation OPACs. In these preliminary notes, the authors point out that in spite of significant enhancements to the user interface and the overall quality of interaction with the OPAC systems - most notably hypertext search and browse features, many previously documented problems experienced by users of first and second generation OPACs are still encountered in Web OPACs. They attribute this to the fact that the underlying structure of the Web OPACs remains unchanged (e.g., record format and content, indexing, and search algorithms).

Project Methodology

Overview

To address the research questions articulated in this study, and to test several research hypotheses, an experiment was designed to compare university students' use and perceptions of two on-line library catalogs (OPACs), a Web OPAC and the conventional version of the same OPAC. A randomized, multi-factor block design was used in this experiment. Four independent test groups were defined by the combination of the two experimental factors, OPAC interface style (Web-based OPAC and Text-based OPAC), and level of search task difficulty (easy and hard - See Table 1). The random assignment of volunteer subjects to these experimental blocks minimized variations between test subjects and testing conditions. Furthermore, this design permits either independent factor, interface style or search task level of difficulty, to be considered as the "treatment" factor. The operational on-line catalog at the University of Tulsa was used for this experiment. The two OPACs differed only in user interface style and interaction capabilities. One version of the OPAC employed the conventional text-based command and menu-driven interface. The other version, the Web OPAC, featured a "point and click" graphical user interface (GUI). Hypertext searching and browsing was supported in the Web version. Both OPACs contained the same catalog database and searchable indexes. Both versions of the OPAC were accessible via the Internet. The Web OPAC was available on the World Wide Web, and the text-based OPAC was available via Telnet access on the Internet.

Participants in the study were recruited from the general undergraduate population of the University of Oklahoma. As volunteers arrived for their scheduled one-hour appointments, they were assigned serially to one of the four test groups, and this continued until a total of 16 subjects completed the test in each of the four experiment groups. The session monitor, either a research assistant or the principle investigator, welcomed the student to the study and the on-line catalog workstation. The workstation was already logged on to the appropriate OPAC. To minimize distractions during the search sessions, the workstation was located in a quiet isolated area of the university library building. No training in the use of the test OPAC was given to the participant. The monitor handed the student participant a clipboard containing the search task worksheet and the post-search questionnaire. After a brief introduction to the project, the monitor left the room but remained on call nearby to assist with any technical problems that might arise. In only one case was it necessary to assist the participant to log back on to the appropriate OPAC. No training in the use of the test OPAC was given to the participant. Most of the 64 participants

completed the search tasks within the allotted one hour. A few were given additional minutes to finish a search in progress and to complete the questionnaire. The session monitor assured each participant that there was no need to rush through the post-search questionnaire.

Data Collection Instruments

During the test session, each participant was required to perform a set of pre-selected search tasks, and to record the results of the searches on the search worksheet. One worksheet contained a set of easy searches (See Appendix A) and another contained a set of hard searches (Appendix B). With assistance from professional librarians and experienced database searchers, two sets of OPAC search tasks were designed, a set of easy searches, and a set of hard searches. The searches were selected for the easy set, and four searches for the more difficult set. With two different OPACs and these two sets of search tasks, four experimental groups were defined by a unique combination of OPAC and search task level of difficulty. For purposes of simplification in the remainder of this report, each of the four test groups will be labeled in accord with its OPAC and task level. Group 1 will be "Text/Easy," Group 2 – "Web/Easy," Group 3 – "Text/Hard," and Group 4 – "Web/Hard." Each participant was instructed to report the results of a search by recording the found item's exact call number on the search worksheet. These call numbers were used later to score the participant's actual search success for data analysis and group comparisons.

After the searches were completed, the participant was required to complete a post-search questionnaire (Appendix C). Only after completing the questionnaire would the participants receive their ten dollar payment. This may explain why there were no missing values on these instruments. The post-search questionnaire consisted of 17 questions (See Appendix C). Questions 1 through 11 were designed to investigate users' perceptions of system ease of use and usefulness, and users' level of satisfaction with search results. Various dimensions of ease of use and searcher satisfaction were addressed by eleven statements. Participants were asked to respond to these statements on 4-point or 5-point Likert scales. Questions 1 through 7 were designed to create an ease of use measurement index. Six additional questions solicited data about the participant's age, university status, major area of study, and previous experience with OPACs and the Web. The gender of each participant was recorded on the questionnaire by the researchers.

The experiment took place over a period of four weeks. A few participants completed their searches and questionnaires in less than one hour. Those who had not finished searching after one hour were given a few more minutes to complete a search in progress. All 64 participants made some attempt to perform each of the searches in their designated search task set. All 64 participants completed the post-search questionnaire. It can be said with some confidence that these students were sincere, motivated, and needy.

Independent and Dependent Variables

In summary, this experiment focused on the following variables: two independent variables, interface style (a WebOPAC and a TextOPAC) and search task level of difficulty, and four dependent variables, actual search performance, perceived ease of use, perceived system usefulness, and perceived search success (i.e., expressed satisfaction with search results).

Research Hypotheses

The following are the research hypotheses tested in this study. As a convention in this report, the two catalogs will be referred to as the "WebOPAC" and the "TextOPAC."

1. Search performance by users of the WebOPAC will be significantly superior to the search performance of users of the TextOPAC.
2. The search performance of users who perform easy search tasks will be superior to the search performance of users who perform difficult searches on both the TextOPAC and the WebOPAC.
3. Searchers' actual search performance will be reflected in their level of satisfaction with search results.
4. More than actual search success and performance, factors such as perceived ease of use and perceived system usefulness will influence searchers' satisfaction with search results.
5. Users will judge the WebOPAC as superior in ease of use.
6. Users will judge the WebOPAC as superior in usefulness, that is, superior in assisting the searcher in finding desired items.

7. Satisfaction with search results will be greater among users of the WebOPAC.
8. Search task level of difficulty will be a significant determinant of users' assessments of the system's ease of use.
9. Search task level of difficulty will be a significant determinant of users' level of satisfaction with search results.

Data Analysis and Findings

Personal Characteristics

The data collected on the personal characteristics of the 64 participants in the study are displayed by group in Tables 2a and 2b. With the exception of gender, these data were collected from participants' answers to the last six questions on the questionnaire. For four of these variables, university status, number of OPACs used, catalog use frequency, and Web searching frequency, original responses choices were consolidated for purposes of data analysis. For example, "University Status" was collapsed from the four categories, Freshman, Sophomore, Junior, and Senior, to lower division and upper division. The number of responses were combined, accordingly. Five response categories for "Catalog Use Frequency" were collapsed into just two, Low and High (High = "Once or Twice a Week" plus "Daily or Almost Daily").

	Gender F M	Mean Age	Univ Status L U	Major H N S
Group 1	11 5	23.4	2 14	3 8 5
Group 2	8 8	20.8	5 11	4 9 3
Group 3	8 8	22.3	6 10	1 12 3
Group 4	8 8	21.4	4 12	4 8 4
Totals (Percent)	35 29 54.7 45.3	21.97	17 47 26.6 73.4	12 37 15 18.8 57.8 23.4

Table 2a. User Profile-A By Group: Personal Characteristics
(Note: Univ. Status *L*=Freshmen/Sophomore; *U*=Junior/Senior
Major *H*=Humanities, *N*=Natural Sciences, *S*=Social Sciences)

	OPACs Used Few Many	Catalog Use Freq Low High	Web Use Freq Low High
Group 1	3 13	10 6	0 16
Group 2	3 13	14 2	1 15
Group 3	4 12	11 5	1 15
Group 4	5 11	11 5	1 15
Totals (Percent)	3 61 4.7 95.3	46 18 71.9 28.1	3 61 4.7 95.3

Table 2b. User Profile-B By Group: on-line Search Experience

Several pertinent facts are revealed in this profile of participants in the study. Most participants were upper division undergraduate students (73.4 %), and most were majoring in one of the natural sciences or mathematics (57.8 %). With regard to search experience, most had used at least 2-3 different library computer catalogs, but most of these reported low frequency of library catalog use (71.9%). On the other hand, all but three of the participants reported high use of a Web search engine or search service to look for information on the Internet.

Search Performance Scoring

Participants were required to record the call numbers of retrieved items on the search task answer sheet. The researchers conducted extensive and repeated searches of the Tulsa University database to discover and retrieve all the relevant or possibly relevant records for each test search. If an item recorded on a participant's answer sheet was not among these records (a rare occurrence), the full bibliographic record was retrieved to assess the relevance or non-relevance of the item.

Each of these found items was judged by the researchers on a 3-level scale of relevant, possibly relevant, and not relevant. Relevant items received 2 points, possibly relevant items received 1 point, and non-relevant items received no points. Each searcher's points were totaled and then transformed to an equivalent number on a 100 point scale. This transformation adjusted for the different amounts of total points achievable on the hard search task sheet (20), as opposed to the easy search task sheet (16). This adjustment made it possible to meaningfully compare the mean search scores of all four groups. The mean search performance scores of each of the four groups are presented in Table 3.

		SEARCH TASKS	
		Easy	Hard
I N T E R F A C E	Text/Menu	Group-1 81.641	Group-3 70.625
	GUI/Web	Group-2 83.984	Group-4 75.625

Table 3. Group Mean Values: Search Performance

Post-Search Questionnaire Responses

Several of the questions on the post-search questionnaire were designed to gather data on users' satisfaction with their search results, and assessments of system ease of use and usefulness. In question 10, "How satisfied were you with the results of your searches?" participants responded on a five-point Likert scale, ranging from "Very satisfied" to "Very unsatisfied." If one assumes a roughly equal distance between each of the values on this five-point scale, the mean values of the four groups shown in Table 4 indicate little or no difference in searchers' level of satisfaction with their search results. (For purposes of analysis, this ordinal scale was transformed into an interval scale as follows: Very satisfied = 20 points, Satisfied = 15 points, and so on.) A closer look at the responses reveals that 37.6 percent of the Group 3 (Text/Hard) participants expressed no more than a "3" level of satisfaction (on the five-point ordinal scale) with their search results.

		SEARCH TASKS	
		Easy	Hard
I		Group-1	Group-3

N T E R F A C E	Text/Menu	13.750	13.750
	GUI/Web	Group-2 15.313	Group-4 13.750

Table 4. Group Mean Values: Search Results Satisfaction Scale

Questions 1-7 were designed to measure several specific aspects of OPAC system ease of use. Question 8 was designed to measure overall ease of use. Each of the questions 1-7 required participants to indicate their level of agreement on a 4-point scale with a statement that described some aspect of interaction with the OPAC. A 4-point scale was used to eliminate fence-sitters. Question 8 asked subjects to assess the "overall ease of use" of the OPAC on a 5-point scale, from "Very difficult" to "Very easy."

It has become common practice in social sciences research to construct multi-factor measurement indices by combining several questions that address related aspects of a single variable (e.g., ease of use). (Sirkin, 1995, 68). While each question may use an ordinal response scale, taken together, the responses to these questions are transformed into a single composite interval level of measurement index. As Schutt explains, "When several questions are used to measure one concept, the responses may be combined by taking the sum or average of responses. A composite measure based on this type of sum or average is termed an index or scale." (Schutt, 1999, 75). And Schutt continues: "In addition, the index can be considered a more complete measure of the concept than can any one of the component questions."

Questions 1-7 were combined into a single interval level index to provide a better, more accurate measure of a participant's assessment of the OPAC's ease of use. Each "Strongly agree" response was assigned 4 points, each "Agree" response 3 points, each "Disagree" response 2 points and so on. A score total of 28 would reflect the strongest agreement on all seven questions. The results of this indexed ease of use measure are provided in Table 5. The responses to question 8, "overall ease of use" appear to support the validity of this index.

		SEARCH TASKS	
		Easy	Hard
I N T E R F A C E	Text/Menu	Group-1 22.250	Group-3 20.250
	GUI/Web	Group-2 23.250	Group-4 22.313

Table 5. Group Mean Values: Ease of Use Index

Question 9 asked the participants to express their level of satisfaction on a 5-point scale with the system's ability "to assist you in finding books." Assuming a roughly equal distance between each of the points on the response scale, the mean levels of "usefulness" ratings for each group are presented in Table 6. The detailed frequency counts for responses to this question reveal that nearly one-half of the Group 3 (Text/Hard) searchers rate the usefulness of the system at no more than level 3 on the 5-point scale. Nearly all Group 2 (Web/Easy) searchers (93.8%) rated the usefulness of their OPAC at a 4 or 5 level.

		SEARCH TASKS	
		Easy	Hard

I N T E R F A C E	Text/Menu	Group-1 3.813	Group-3 3.625
	GUI/Web	Group-2 4.125	Group-4 3.938

Table 6. Group Mean Values: Usefulness Satisfaction Rating

Independent Group Comparisons

Table 7 displays the mean values of the four dependent variables by test group. To compare the mean search performance scores and rating assessments of the four groups, and to test the research hypotheses, both parametric (independent group, two-sample t-tests) and non-parametric (Mann-Whitney tests) statistical procedures were employed in the analysis of the data presented here. When the data are not normally distributed, the Mann-Whitney tests produce more valid results.

	Mean Search Scores	Mean Results Satisfaction	Mean Ease of Use Rating	Mean Usefulness Rating
1-Text/Easy	81.6406	13.7500	22.2500	3.8125
2-Web/Easy	83.9844	15.3130	23.2500	4.1250
3-Text/Hard	70.6250	13.7500	20.2500	3.6250
4-Web/Hard	75.6250	13.7500	22.3125	3.9375

Table 7. Original Four Test Group Comparisons

These comparisons of each pair of the test groups were conducted to discover any statistically significant differences in the values of the four dependent variables. Such differences would support the rejection of one or more null hypotheses (no difference) regarding these variables, and provide support for the research hypothesis.

Search Performance Scores

Each group's search performance scores were compared to every other group's scores. Of the six possible pairings, significant differences were found between Group 1 and Group 3 ($p=0.024$), as well as between Group 2 and Group 3 ($p=0.006$). (In all these comparisons, the level of significance chosen is 0.05.) The Mann-Whitney tests support these findings, and also suggest a significant difference exists between the performance scores of Group 2 and Group 4 ($p=0.037$). WebOPAC searchers significantly outperformed the TextOPAC searchers, but easy task searchers also outperformed hard task searchers on both OPACs.

Satisfaction With Search results

After a glance at Table 7, one might conclude that there was no appreciable difference among searchers in the four groups when it came to expressing satisfaction with their search results. The results of the t-tests and Mann-Whitney tests confirm this conclusion. No significant differences were identified. This suggests a disjunction between actual search results and subjective assessments of the quality of those results. Those participants with the lowest search

performance scores express equal or nearly-equal levels of satisfaction with search results as those with much higher search scores.

Ease of Use Assessments

Questions 1-7 on the post-search questionnaire comprised the ease of use index which made possible an interval level of measurement. The mean rating scores of the four test groups are displayed in Table 7. The most significant difference in user assessments of OPAC ease of use exists between Group 2 and Group 3 ($p=0.010$). A significant difference in ease of use assessments was also discovered between Group 3 and Group 4 ($p=0.040$). Mann-Whitney tests strongly supported the t-test results for Groups 2 and 3 ($p=0.006$), but provided only weak support for the Groups 3 and 4 finding ($p=0.068$). Both WebOPACs received higher ease of use ratings than their text counterparts.

OPAC Usefulness Assessments

Question 9 on the questionnaire was designed to measure searcher's satisfaction with "the ability of the library computer catalog to assist you in finding the books you were looking for." Participant's responses were measured on a 5-point satisfaction scale, with 5 points for "Very satisfied." In future research a multi-question "usefulness" index might be constructed to enable more refined measures of this important variable. Based on responses to this single question, no significant differences between groups were discovered using either the t-test or the Mann-Whitney test. One cannot fail to observe, however, that Group 2 (Web/Easy) users rated their OPAC higher in usefulness than Group 3 (Text/Hard) users. It is not clear at this point whether this can be attributed to the interface or to the task level.

Type of OPAC and Task Level Group Comparisons

The four original test groups can be realigned to create four new groups for purposes of analysis. The four new groups include, respectively, all participants who used the TextOPACs, all those who used the WebOPACs, all those who performed the easy search tasks, and all those who performed the hard search tasks. The mean values for these groups, so considered, are displayed in Table 8.

	Mean Search Scores	Mean Results Satisfaction	Mean Ease of Use Rating	Mean Usefulness Rating
Text OPACs	76.1328	13.7500	21.2500	3.7188
Web OPACs	79.8047	14.5313	22.7813	4.0313
Easy Tasks	82.8125	14.5313	22.7500	3.9688
Hard Tasks	73.1250	13.7500	21.2813	3.7813

Table 8. Group Comparisons By OPAC and Task Level

Immediately it can be seen that the highest level of search performance was achieved on the easy search tasks, while the lowest level of performance was registered by those required to carry out the hard searches. WebOPACs received higher ease of use and usefulness ratings than TextOPACs. The superiority in search performance achieved by the easy task searchers over the hard task searchers is significant ($p=0.005$). WebOPAC searchers' performance scores were higher than the performance scores of TextOPAC searchers, but this difference was not found to be significant.

Assessments of ease of use and system usefulness appear to be affected by search task level of difficulty. Easy task searchers rated their OPACs higher in assessments of ease of use and system usefulness. The higher ease of use rating given by the easy task searchers was statistically significant (t-tests: $p=0.05$; Mann-Whitney: $p=0.033$).

Additional t-test analysis demonstrated that the ease of use ratings provided by the TextOPAC users and the WebOPAC users are significantly different ($p=0.041$). The WebOPAC was rated superior in ease of use. The WebOPAC usefulness ratings were higher than the TextOPAC usefulness ratings, but this difference was not statistically significant. Regarding satisfaction with search results, no significant group differences were identified.

Associations Between Dependent Variables

Pearson's correlations coefficient (r), a parametric procedure, and Spearman's Rank correlation coefficient, a non-parametric procedure, were used to examine the relationships between the four dependent variables, interpreted as quantitative variables. No significant associations were found between search performance scores and any of the other three dependent variables (ease of use, usefulness, and satisfaction with search results). A strong positive correlation was identified between the satisfaction with search results ratings and the ease of use assessments ($r=0.5992$, $p<0.001$). A strong positive correlation was also identified between satisfaction with search results and perceived system usefulness ($r=0.7113$, $p<0.001$). Furthermore, a strong positive association exists between searchers' ease of use ratings and their satisfaction with the system's usefulness in assisting them in finding relevant items ($r=0.7080$, $p<0.001$). This analysis indicates that actual search performance is not a predictor of a user's satisfaction with search results. It does indicate that users' perceptions of ease of use and system usefulness exert a stronger influence on satisfaction with results than actual search performance.

Tests for Independence

Tests for independence (Chi Square) were conducted to examine whether participants grouped by type of OPAC used or by level of search task difficulty responded differently to questions 1-9 (interpreted as category responses) on the post-search questionnaire. This analysis revealed a significant association between participants' responses to question 5, "My parents could use this library computer catalog to search effectively with little or no training", and the type of OPAC used ($p=0.005$). A significantly greater number of WebOPAC users agreed with this statement than did TextOPAC users (See Table 9).

FREQUENCY: TOT PERCENTAGE: ROW PERCENTAGE: COL PERCENTAGE:	Disagree	Agree	TOTAL
Text OPAC	8 16.7 50.0 66.7	8 16.7 50.0 22.2	16 33.3
Web OPAC	4 8.3 12.5 33.3	28 58.3 87.5 77.8	32 66.7
TOTALS	12 25.0	36 75.0	48 100.0

Table 9. "My parents could use this OPAC..." (Question 5)

WebOPAC users seem to be saying, "WebOPACs are so easy to use, even my parents could use them effectively with little or no training!" An association was also identified between type of OPAC used and responses to question 9, "How satisfied were you with the ability of the library computer catalog to assist you in finding the books you were looking for?" ($p=0.016$). WebOPAC users were overwhelmingly satisfied with the usefulness of their OPACs, TextOPAC users less so (See Table 10).

FREQUENCY: TOT PERCENTAGE:			
-------------------------------	--	--	--

ROW PERCENTAGE: COL PERCENTAGE:	Not Satisfied	Satisfied	TOTAL
Text OPAC	11 17.2 34.4 78.6	21 32.8 65.6 42.0	32 50.0
Web OPAC	3 4.7 9.4 21.4	29 45.3 90.6 58.0	32 50.0
TOTALS	14 21.9	50 78.1	64 100.0

Table 10. System Usefulness Results (Question 9)

Associations Between Personal Characteristics and the Dependent Variables

Chi Square tests for independence were conducted to identify any significant associations between the personal characteristics and any of three dependent variables, satisfaction with search results, ease of use, and system usefulness. No associations were found between any of these variables and gender, subject major, number of OPACs used, or catalog use frequency. An association was discovered between university status and ease of use ($p=0.010$). Upper division students rated their OPACs ease of use much higher than lower division students rated theirs. (See Table 11).

FREQUENCY: TOT PERCENT: ROW PERCENT: COL PERCENT:	2	3	4	5	TOTAL
Lower Division	3 4.7 17.6 100.0	1 1.6 5.9 20.0	6 9.4 35.3 16.2	7 10.9 41.2 36.8	17 26.6
Upper Division	0 .0 .0 .0	4 6.3 8.5 80.0	31 48.8 66.0 83.8	12 18.8 25.5 63.2	47 73.4
TOTALS	3 4.7	5 7.8	37 57.8	19 29.7	64 100.0

Table 11. Ease of Use ratings By University Status (5=Very Easy)

Associations were also discovered between frequency of Web searching and satisfaction with search results ($p=0.033$), and between frequency of Web searching and assessments of system usefulness ($p=0.048$). Frequent Web searchers expressed higher levels of satisfaction with search results and system usefulness than infrequent Web searchers.

Summary and Conclusion

The discussion of the findings in this study focused on three areas: interactions between the independent variables (interface style and task level of difficulty) and the four dependent variables (search performance, satisfaction with

search results, ease of use, and usefulness); associations between dependent variables; and associations between personal characteristics and dependent variables. This research indicates that both interface style and search task level of difficulty exert an influence on search performance and ease of use or usefulness factors. Performing the same search tasks, WebOPAC users scored higher than TextOPAC users, but the difference was not statistically significant at the .05 significance level. Using the same OPAC, easy task searchers outperformed hard task searchers by a wide margin. The difference was found to be significant. The magnitude of this difference suggests that task level of difficulty exerted a stronger influence on performance scores than interface style.

WebOPACs were rated higher than TextOPACs in ease of use and usefulness. The difference was found to be significant. Easy task searchers rated their system's ease of use higher than hard task searchers, but this difference is not significant. No significant differences were found between any of the groups in participants' satisfaction with search results or assessments of system usefulness.

Few associations were found between personal characteristics and the dependent variables. Upper division students rated both OPACs easier to use than lower division students. A significant association was discovered between frequency of Web searching and two variables, satisfaction with results, and system usefulness. Frequent Web searchers expressed much higher levels of satisfaction with search results and system usefulness. They also were more likely to use WebOPACs again than infrequent Web searchers. Web use seems to lead to more Web use.

The results of the correlation analysis are the most telling. Once again, the evidence indicates that actual search performance is not a predictor or determinant of a searcher's satisfaction with search results. There seems to be little interaction between these two variables. On the other hand, perceptions of ease of use and system usefulness during the search process do influence users' satisfaction with search results.

In summary, when considering search performance, search task level of difficulty seems to be a major determinant, but OPAC interface style may affect search performance as well. When considering perceptions of ease of use and usefulness, interface style appears to be the primary determinant. Furthermore, as earlier studies have shown, there is little or no association between actual search performance by users and their expressed satisfaction with search results. In short, WebOPACs are easier to use, and this may be a supportive factor in search success. However, the down side is this: WebOPACs may contribute to "false positives" in users' assessments of search results, and this may explain, at least partially, why users are often satisfied with poor search results.

Research Hypotheses Revisited

In light of these findings, each of the ten research hypotheses stated earlier in this report can now be reviewed. Which of them is supported by the evidence, which may be supported, and which are not supported?

1. Search performance by users of the WebOPAC will be significantly superior to the search performance of users of the TextOPAC.

Supported?: Maybe. WebOPAC search performance scores are higher, but this difference is not significant at the .05 level.

2. The search performance of users who perform easy search tasks will be superior to the search performance of users who perform difficult searches on both the TextOPAC and the WebOPAC.

Supported?: Yes. This difference is significant ($p=0.005$).

3. Searchers actual search performance will be reflected in their level of satisfaction with search results

Supported?: No. There appears to be little interaction between these two variables. Searchers may express satisfaction with search results even when the results are far from optimal.

4. More than actual search success and performance, factors such as perceived ease of use and perceived system usefulness will influence searchers' satisfaction with search results.

Supported?: Yes. Strong positive correlations were identified between ease of use assessments and satisfaction with search results. Strong positive correlations were identified between perceived system

usefulness and satisfaction with search results.

5. Users will judge the WebOPAC as superior in ease of use.

Supported?: Yes. The data show this difference to be significant ($p=0.041$).

6. Users will judge the WebOPAC as superior in usefulness, that is, superior in assisting the searcher in finding desired items.

Supported?: Maybe. WebOPAC usefulness ratings are higher than the TextOPAC usefulness ratings, but this difference is not statistically significant at the .05 level.

7. Satisfaction with search results will be greater among users of the WebOPAC.

Supported?: No. There is not sufficient evidence to support this. However, a strong positive association was found between ease of use and satisfaction with search results.

8. Search task level of difficulty will be a significant determinant of users' assessments of the system's ease of use.

Supported?: Yes. Task level of difficulty has an affect on users' assessments of ease of use. The easy task searchers rated their OPACs significantly higher in assessments of ease of use

9. Search task level of difficulty will be a significant determinant of users' level of satisfaction with search results.

Supported?: No. Easy task searchers rated their OPACs higher in assessments of usefulness, but the difference was not statistically significant.

Reflections on the Study

Participants selected themselves to participate in this study. For this reason, caution must be exercised in extending these findings to any larger population of OPAC users. The external validity of the study has not been established. Participants were randomly assigned to the test groups, but this aggregate of volunteer participants constitutes a convenience sample, not a true probability sample. The findings of this study may not be representative of any other community of OPAC users.

Assigning numerical scores to search results is always a risky business. Some of the problems associated with recall and precision scoring were avoided in this study, but the scoring of results was based on "relevance" judgments made by the searchers and the researchers. These judgments were made after viewing only standard catalog records, well-known for their paucity of "content" information that is helpful in such judgments. The scoring of search results was done as consistently as possible, and the total score of each searcher was derived from the items found and recorded. Relevant items they did not retrieve did not factor into these computations. Nonetheless, this searching took place in an "artificial" setting where a searcher's actual information needs and search objectives were not factored into the search attempts or the relevance assessments. Judgments of relevance made in such settings are understandably open to question.

A multi-question index was designed to measure users' perceptions of system ease of use. A similar index for measuring perceptions of system usefulness in assisting the searcher was not developed. This should be attempted in future studies. If these two variables, ease of use and system usefulness, are different, then great care must be exercised in devising separate measurement indexes for each. The validity and reliability of these indexes must be established.

Researchers must continue to devise refined measures of user satisfaction with search results and system performance. Satisfaction with search results is often influenced by non-performance factors. More testing and evaluation of satisfaction measurement indexes is needed. In the meantime, satisfaction as a measure of retrieval effectiveness must remain highly suspect. It may be the case, however, that we have reached the limits of quantitative research methods in attempts to gain an understanding of false positives in users' assessments of IR system performance and effectiveness. Qualitative research methods may assist us in achieving a better

understanding of this complex phenomenon of user satisfaction with the output and performance of information retrieval systems.

References

- Adams, Dennis A., Nelson, R. Ryan, and Todd, Peter A. (1992). Perceived usefulness, ease of use, and usage of information technology: a replication. *MIS Quarterly*, June 1992. 227-247.
- Ankeny, Melvon L. (1991). Evaluating end-user services: success or satisfaction? *The Journal of Academic Librarianship*, 16(6): 352-356.
- Applegate, Rachel. (1993). Models of user satisfaction: understanding false positives. *RQ*, 32(4): 525-539.
- Ayres, F.H., Nielsen, L.P.S., and Ridley, M.J. (1998). *The Bradford OPAC 2 (BOPAC2): managing and displaying retrievals from a distributed search in Z39.50*. London, UK: British Library Research and Innovation Centre. (British Library Research and Innovation Report 103)
[<http://www.bopac2.comp.ac.uk/~bopac2/report/>]
- Bates, Marcia J. (1990). Where should the person stop and the information search interface start? *Information Processing and Management*, 26: 575-591.
- Beaulieu, Micheline, and Borgman, Christine L. (1996). A new era for OPAC research: introduction to special topic issue on current research in on-line public access systems. *Journal of the American Society for Information Science*, 47(7): 491-492.
- Belkin, Nicholas J. and Vickery, Alina. (1985). *Interaction in information systems: a review of research from document retrieval to knowledge-based systems*. London, UK: British Library. (Library and Information Research Report , no. 35).
- Borgman, Christine L. (1996). Why are on-line catalogs still so hard to use? *Journal of the American Society for Information Science*, 47(7): 493-503.
- Chen, Hsinchun and Dhar, Vasant. (1990). User misconceptions of information retrieval systems. *International Journal of Man-Machine Studies*, 32: 673-692.
- Cherry, Joan M. (1997). Bibliographic displays in OPACs and Web catalogs: how well do they comply with display guidelines? *Information Technology and Libraries*, 17(3):
- Cherry, J.M. and Cox, P.J. (1996). World Wide Web displays of bibliographic records: an evaluation. In *Proceedings of the 24th Annual Conference of the Canadian Association for Information Science*. Banff, Canada: Canadian Association for Information Science. 101-114.
- Cleverdon, Cyril W. (1970). *The effect of variations in relevance assessments in comparative experimental tests of index languages*. Cranfield, UK: Cranfield Institute of Technology. (Cranfield Library Report No. 3)
- Cleverdon, Cyril W., Mills, Jack, and Keen, E. Michael. (1966). Factors determining the performance of indexing systems. Cranfield, UK: Aslib Cranfield Research Project, College of Aeronautics. (Volume 1: Design; Volume 2: Results)
- Cooper, William S. (1973). On selecting a measure of retrieval effectiveness. Part I. The "subjective" philosophy of evaluation; Part II. Implementation of the philosophy. *Journal of the American Society for Information Science*, 24: 87-100; 413-424.
- Crawford, Gregory A., Lee, Arnold, Connolly, Lorene, and Shylaja, Y.L. (1992). OPAC user satisfaction and success: a study of four libraries. In *7th Conference on Integrated on-line Library Systems, IOLS 1992: Integrated on-line Library Systems*. New York, NY: Learned Information.
- Davis, Fred D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, September 1989. 319-340.
- Delone, William H, McLean, Ephraim R. (1992). Information systems success: the quest for the dependent variable. *Information Systems Research*, 3: 60-95.
- Dervin, Brenda and Nilan, Michael S. (1986). Information needs and uses. In Williams, Martha E. ed. *Annual Review of Information Science and Technology (ARIST)*, 21: 1986. White Plains, NY: Knowledge Industry Publications, Inc. 3-33.
- Ellis, David. (1992). The physical and cognitive paradigms in information retrieval research. *Journal of Documentation*, 48: 45-64.
- Ellis, David. (1994). Paradigms in information retrieval research. In Kent, Allen, ed. *Encyclopedia of Library and Information Science*, Volume 54: 275-291. New York, NY: Marcel Dekker, Inc.
- Ellis, David. (1996). *Progress and Problems in Information Retrieval*. 2nd edition. London, UK: Library Association Publishing.
- Gatian, Amy W. (1994). Is user satisfaction a valid measure of system effectiveness? *Information and*

Management, 26: 119-131.

- Gluck, Myke. (1996). Exploring the relationship between user satisfaction and relevance in information systems. *Information Processing and Management*, 32(1): 89-104.
- Harmon, Donna K. (1992). Evaluation issues in information retrieval. *information processing and management*, 28(4):439-440.
- Harmon, Donna K., ed. (1995). *Overview of the 3rd Text Retrieval Conference (TREC-3); November 2-4, 1994*, Gaithersburg, MD. Gaithersburg, MD: National Institute of Standards and Technology. (NIST Special Publication 500-225).
- Harmon, Donna K., ed. (1996). *Overview of the 4th Text Retrieval Conference (TREC-4); November 1-3, 1995*, Gaithersburg, MD. Gaithersburg, MD: National Institute of Standards and Technology. (NIST Special Publication 500-236).
- Harter, Stephen P. and Hert, Carol A. (1997). Evaluation of information retrieval systems: approaches, issues, and methods. In Williams, Martha E. ed. *Annual Review of Information Science and Technology (ARIST)*, 32: 1997. Medford, NJ: Information Today. 3-94.
- Hildreth, Charles R. (1995a). on-line catalog design models: are we moving in the right direction? A report commissioned by the *Council on Library Resources*, August 1995. <http://phoenix.liunet.edu/~hildreth/clr-opac.html>.
- Hildreth, Charles R. (1995b). The GUI OPAC: approach with caution. *The Public Access Computer Systems Review*, 6(5): 1995. (<http://info.lib.uh.edu/pr/v6/n5/hild6n5.html>)
- Hildreth, Charles R. (1997). The use and understanding of keyword searching in a university on-line catalog. *Information Technology and Libraries*, 16(2): June 1997. 52-62.
- Hiltz, Starr Roxanne and Johnson, Kenneth. (1989). Measuring acceptance of computer-mediated communications systems. *Journal of the American Society for Information Science*, 40 (6): 386-397.
- Lancaster, F.W., Elzy, Cheryl, Zeter, Mary Jo, Metzler, Laura, and Low, Yuen-Man. (1994). Searching databases on CD-ROM: comparison of the results of end-user searching with results from two modes of searching by skilled intermediaries. *RQ*, 33(3): 370-386.
- Large, Andrew and Beheshti, Jamshid. (1997). OPACs: A Research Review. *Library & Information Science Research*, 19(2): 111-133.
- Mizzaro, S. (1997). Relevance: the whole history. *Journal of the American Society for Information Science*, 48(9): 810-832.
- Nahl, D. and Meer, M.P. (1997). User-centered assessment of two Web browsers: errors, perceived self-efficacy, and success. In *Proceedings of the American Society for Information Science (ASIS) 60th Annual Meeting*: Volume 34; 1997. Medford, NJ: Information Today, Inc. 89-97.
- Nitecki, Danuta. (1993). User criteria for evaluating the effectiveness of the on-line catalog. In Murfin, Marjorie E. and Whitlatch, Jo Bell, eds., *Research in Reference Effectiveness, RASD Occasional Papers*, Number 16. Chicago: Reference and Adult Services Division, American Library Association, 1993. 8-28.
- Ortiz-Repiso, Virginia, and Moscoso, Purificacion. (1999). Web-based OPACs: between tradition and innovation. *Information Technology and Libraries*, June 1999: 68-77.
- Sandore, Beth (1990). on-line searching: what measure satisfaction? *Library and Information Science Research*, 12: 33-54.
- Saracevic, Tefko, and Kantor, Paul. (1988). A study of information seeking and retrieving. II. users, questions, and effectiveness. *Journal of the American Society for Information Science*, 39(5): 177-196.
- Schutt, Russell K. (1999). *Investigating the social world: the process and practice of research*, 2nd ed. Thousand Oaks, CA: Pine Forge Press.
- Sirkin, Mark R. (1995). *Statistics for the social sciences*. Thousand Oaks, CA: Sage Publications.
- Soergel, Dagobert. (1976). Is user satisfaction a hobgoblin? *Journal of the American Society for Information Science*, 27(July-August, 1976): 256-259.
- Steffey, Ramona J. and Meyer, Nikki. (1989). Evaluating user success and satisfaction with CD-Rom. *Laserdisk Professional*, September 1989. 35-44.
- Su, Louise T. (1992). Evaluation measures for interactive information retrieval. *Information Processing and Management*, 28: 503-516.
- Su, Louise T. (1994). The relevance of recall and precision in user evaluation. *Journal of the American Society for Information Science*, 45(3): 207-215.
- Su, Louise T. (1996). Value of search results as a whole as a measure of information retrieval performance. In Hardin, Steve, ed: *Proceedings of the American Society for Information Science (ASIS) 59th Annual Meeting*: Volume 33; 1996 October 21-24. Baltimore, MD. Medford, NJ: Information Today, Inc. 226-237.

Tessier, Judith A., Crouch, Wayne W., Atherton, Pauline. (1977). New measures of user satisfaction with computer-based literature searches. *Special Libraries*, 68(November 1977): 383-389.

- Tinanoff, Susan. (1996). End-user searching with CD-ROM databases. *The Katherine Sharp Review*, 2 (Winter 1996). [<http://edfu.lis.uiuc.edu/review/winter1996/tianoff.html>]
 - Tonta, Yasar. (1992). Analysis of search failures in document retrieval systems: a review. *The Public Access Computer Systems Review*, 3(1): 4-53.
-

Appendix A (Easy search task worksheet)

LIBRARY COMPUTER CATALOG RESEARCH PROJECT

Thank you for volunteering to participate in this study. You and your contributions will be kept strictly anonymous. In the next hour you will be required to search for library books in an online library computer catalog. After you have finished searching, you will be required to answer a few questions about your catalog search experiences on the Post-Search Questionnaire. Please try to be as sincere as possible in your search efforts and responses to the Post-Search Questionnaire. Your contributions may lead to enhancements to today's library computer catalogs, by providing much-needed knowledge about their use and usefulness.

INSTRUCTIONS:

Please complete the following three searches for books in this catalog only. When you find a book you think is appropriate, write its call number in the space designated. The call number will appear in a longer, detailed display of a single book's catalog record and will look something like this: **HD6666 .S88 1994**. (Include the date.)

When you are finished with your searches, complete the attached **Post-Search Questionnaire**. THANK YOU!

1. Search for two books by the author, *Drucker, Peter*.

Call number 1: _____

Call number 2: _____

2. Find two books whose titles indicate they are about *sex discrimination in the workplace*.

Call number 1: _____

Call number 2: _____

3. Find a few books on the topic of *economic sanctions in South Africa*.

Call number: _____

Call number: _____

Call number: _____

Call number: _____

Call number: _____

Appendix B (Hard search task worksheet)

LIBRARY COMPUTER CATALOG RESEARCH PROJECT

Thank you for volunteering to participate in this study. You and your contributions will be kept strictly anonymous. In the next hour you will be required to search for library books in an online library computer catalog. After you have finished searching, you will be required to answer a few questions about your catalog search experiences on the Post-Search Questionnaire. Please try to be as sincere as possible in your search efforts and responses to the Post-Search Questionnaire. Your contributions may lead to enhancements to today's library computer catalogs, by providing much-needed knowledge about their use and usefulness.

INSTRUCTIONS:

Please perform the following searches for books in this catalog only. When you find a book you think is appropriate, write its call number in the space designated. The call number will appear in a longer, detailed display of a single book's catalog record and will look something like this: **HD6666 .S88 1994**. (Include the date.)

When you are finished with your searches, complete the attached **Post-Search Questionnaire**. THANK YOU!

1. Find an English translation of a volume of short novels by the French author,

Marie-Henri Beyle.

Call number: _____

2. Find two books published since 1990 on the *treatment of shoulder injuries in sports*.

Call number 1: _____

Call number 2: _____

3. Find a book on *aids and biological warfare*.

Call number: _____

4. Find a few good books on the topic of *prevention of child abuse*.

Call number: _____

Call number: _____

Call number: _____

Call number: _____

Call number: _____

Call number: _____

Appendix C (Post-Search Questionnaire)

Appendix C

Library Computer Catalog Research Project,
Post-Search Questionnaire
(Please circle just one response per question)

1. It was easy to get the library computer catalog to do what I wanted it to do. (Circle one)

Strongly Disagree Disagree Agree Strongly Agree
2. Interaction with the library computer catalog to carry out my searches was clear and understandable.

Strongly Disagree Disagree Agree Strongly Agree
3. Overall, this library computer catalog was helpful in assisting me to search effectively.

Strongly Disagree Disagree Agree Strongly Agree
4. It would take a great deal of effort and practice for me to learn to use this library computer catalog with proficiency.

Strongly Disagree Disagree Agree Strongly Agree
5. My parents could use this library computer catalog to search effectively with little or no training.

Strongly Disagree Disagree Agree Strongly Agree
6. Once I found a book on a subject or topic, it was easy to continue the search to find additional books on the subject.

Strongly Disagree Disagree Agree Strongly Agree
7. When a full, detailed catalog record for a book was displayed, it was easy to read that display to find specific information about the book.

Strongly Disagree Disagree Agree Strongly Agree
8. Overall, how would you rate this library computer catalog in terms of ease of use?
(Circle a single number)

Very Difficult➔ 1 2 3 4 5 ⬅Very Easy

Appendix C

9. How satisfied were you with the ability of the library computer catalog to assist you in finding the books you were looking for?

Very Unsatisfied → 1 2 3 4 5 ← Very Satisfied

10. How satisfied were you with the results of your searches?

Very Unsatisfied → 1 2 3 4 5 ← Very Satisfied

11. There are many library computer catalogs now available on the Internet. How likely would you be to use a library computer catalog like this one in the future?

Not Likely → 1 2 3 4 5 ← Very Likely

#####

Please answer a few questions about yourself:

a. Your age: _____

b. University status: freshman ___ sophomore ___ junior ___ senior ___ other _____

c. Current or planned major: _____

d. How many different library computer catalogs have you used prior to this experiment?

___ None
___ One
___ Two-Three
___ More Than Three

e. Please check the frequency that most accurately represents how often you use a library computer catalog:

___ Never
___ Once or Twice a Year
___ Once or Twice a Month
___ Once or twice a Week
___ Daily or Almost Daily

f. How often have you used an Internet/Web search engine or search service to look for information on the Internet?

___ Never
___ Once or Twice
___ Several Times (3-5)
___ Many Times

Page two (Thank you)

[Information Research, Volume 6 No. 2 January 2001](#)

Accounting for users' inflated assessments of on-line catalogue search performance and usefulness: an experimental study, by [Charles R. Hildreth](#)

Location: <http://InformationR.net/ir/6-2/paper101.html>

© the author, 2001.

Last updated: 19th December 2000
