

# Undergraduate Students in the Digital Library: Information Seeking Behavior in a Heterogeneous Environment

*Peggy Seiden, Kris Szymborski, and Barbara Norelli  
Skidmore College, Saratoga Springs, NY*

## ABSTRACT

---

There is a paucity of research on general undergraduate user behavior particularly in the emerging digital library environment. This paper presents the results of a preliminary study of information seeking among sixty undergraduates at Skidmore College. The study was designed to 1) assess students' information seeking behavior in general; 2) elicit information about the search process; and 3) discern how students had acquired their knowledge of online searching and their level of expertise with online searching, computer applications and libraries. Our results indicate that most undergraduates have a relatively poor understanding of the information environment and that the "digital library" exaggerates and magnifies these problems.

---

## Introduction

In the foreseeable future it is likely that personal information environments will continue to be hybrid of both digital and more traditional library sources. At institutions like Skidmore, the number of digital bibliographic and fulltext resources is rapidly proliferating and playing an ever larger role in students' research process. Therefore, we need to know how users will integrate the digital library into their repertoire of resources, how it will be integrated into their individual information seeking behavior, and how it will impact that behavior. We need to understand how, why and when people use digital resources, especially if we seek to improve the ways we teach users about these resources.

Much of the research on digital library users has up until now focused on faculty and graduate students and their research patterns - eg. Bishop's studies at Illinois<sup>(1)</sup> and where the broader public has been studied it has generally been in reference to very narrow databases such as the Berkeley digital library project<sup>(2)</sup> on water use documents or Bishop's earlier work on visitors to Web based museum sites.<sup>(3)</sup>

But there have been virtually no indepth studies of undergraduates' use of these resources and how they are integrated into their overall information seeking behavior. As the digital library evolves, we hypothesize that their information seeking patterns will evolve - and anecdotal evidence points to the fact those patterns are already changing at libraries like Skidmore's where there are a wide variety of networked bibliographic and fulltext databases available to all students.

## Purpose and Scope of the Research

This paper reports the preliminary findings of an ongoing study of undergraduate information seeking behavior utilizing digital resources at the Scribner Library, Skidmore College. It grew out of a study on faculty research patterns which attempted to discern how liberal arts college faculty's research methods

were altered by new patterns and forms of scholarly communication. That study found that faculty's research patterns remained essentially unchanged with few exceptions, but that they integrated new sources and formats of information into their repertoire in an incremental fashion.(4)

But what impact does this new environment have on the undergraduate student, most of whose research patterns are essentially unformed by disciplinary norms. How does it shape their research patterns?

Many online user studies attempt to ascertain search behavior in single systems such as online catalogs or Dialog databases. While the Skidmore study does consider how students search a specific system of their choice, it does so within the wider framework of trying to understand and model their overall information seeking strategy. To develop this model we gathered a wide variety of data about how students made decisions about which resources to use, the types of search strategies they were likely to employ, the problems they encountered. It is important to note that unlike many information retrieval studies, we did not attempt to measure search performance. Rather we asked users about their expectations for the search outcomes and their perceptions of search success. The study looked at how users said they learned how to search, their prior library and computer experience and why they used database searching as opposed to other research methods.

### **We expected that this data would**

- provide insights into individual's information seeking behavior
- allow us to infer the user's model/understanding of the information environment from his/her behavior.
- tell us what models/behaviors were likely to be more successful in this heterogeneous environment
- provide clues as to what individual differences or other factors might contribute to development of more successful mental models.

It should be noted that a user's mental model of the information environment includes within it his/her model of particular databases as well as his/her model of the environment, its organization and content.

### **Prior Research**

The knowledge base related to user studies is very rich. A cursory search of Library Literature reveals over 1500 articles on the topic, and at least one-tenth which are about information systems and information networks. In the last decade considerable progress has been made in developing a theoretical basis for studying users and modelling users.

There were three specific types of user studies which seemed most applicable to our research: 1) other studies of undergraduate information seeking and library use; and 2) information retrieval studies, particularly those which have studied novice users; and 3) general end user searching studies.

### **Research on Undergraduate Information Seeking and Library Use**

Previous studies on undergraduate information seeking and library use, though not explicitly studying interaction with digital resources, provide some insight into the undergraduate research process. Particularly relevant is Caroline Kuhlthau's information search model, developed from case studies and student papers, which identifies six stages of the information search process: 1) task initiation, 2) topic selection; 3) prefocus exploration, 4) focus formulation, 5) information collection; and 6) search closure and writing initiation. While

Kuhlthau's original study focused on twelfth graders, she tested her model among academic library users and found that it fit well with a few changes - the most significant being a greater emphasis on periodicals among college students.(5)

Building on Kuhlthau's work, Barbara Fister interviewed fourteen research successful undergraduates and found few consistencies in research strategies and no relation to those strategies typically taught in bibliographic instruction classes.(6) Using focus groups and individual interviews, Valentine characterized undergraduates' strategies as motivated by the quick and easy method of doing research.(7)

Both Fister and Kuhlthau emphasize focus formulation as a crucial step in the undergraduate research process.(8) College students will often modify their topic as their research progresses and Fister and Constance Mellon found that writing does not begin after the search closure, but is an integral part of the entire research process and helps in the process of topic development .(9)

## **Information Retrieval Studies**

In the digital library we're not simply interested in how students go about their research, but in how they interact with information retrieval systems, because these systems become the entire information environment in certain problem contexts and for certain users.

Since the mid-1980s there has been considerable amount of research on the search process and factors which impact search outcomes. Fidel and Soergel by studying professional searchers found over 200 variables that could affect search outcome.(10) Saracevic's study of information seeking and retrieving provides a model of the search process and the sets of variables 1) the user and the context of information seeking; 2) the structure and the characteristic of the questions; 3) the searcher, 4) the search, and 5) the items retrieved) which impact on each part of the process. Although his study focused on mediated searching, many of his findings apply to end user searching, as well.(11)

While many classic information retrieval studies have examined search performance by using outcome measures, studies of the search process provide more relevant data. Data about novice searchers is particularly compelling for our purposes. Studies which have looked at both professional and non-professional searchers have found that their searches share key characteristics: they use predominantly search and display commands; they often use AND; they seldom modify their searches; and they conduct brief, simple searches. Furthermore, among novice searchers the use of a thesaurus, even when one is trained in its use, is rare.(12) Meadow et al. has noted that novices tend to learn only the minimal amount necessary to get results from a retrieval system.(13)

While the majority of studies examine citation databases, a growing number of studies focus on fulltext sources. Tenopir and Nahl's research on Magazine Index, ASAP, was one of the first to look at how and why various groups of users searched a general fulltext database.(14) Sievert and McKinin looked at why most fulltext search strategies miss a large number of relevant documents and found that either the strategy was too restrictive or that natural language was insufficient to retrieve all variants of a term or name.(15) Wildemuth's work with novice searchers of a biomedical documents database found that they typically selected a few key concepts to search first and then modified (on average about three times/search) and that the average number of moves decreased as students became more experienced. She also found that there was significant variation in search behavior even among novice searchers.(16)

There are as yet no systematic studies on information seeking on the World Wide Web that examine either why it is used in the context of research or how it is used. The language surrounding discussions of the Web suggests that information seeking on the Web is less retrieval and more "discovery" and that the interface is

conducive to browsing,(17) but whether undergraduates perceive the environment as substantially different from other sources of information is unknown.

There is a significant body of research on how users evaluate systems and their output. The Rutgers study (Saracevic, Spink, Su) of mediated searching looks at types of feedback which are the basis for users' decisions.(18) Taemin Kim Park's work on relevance suggests that the user's perspective on a citation is supported by layers of contextual factors.(19) Kuhlthau's work corroborates Park's theory about the importance of context. She notes that whereas students initially seek "relevant" information related to a general topic, after they focus they seek pertinent information.(20) While relevance is one dimension of how users judge a system, the Rutgers study found 26 facets of a search (only 4 of which pertained to relevance) which influence users' perceptions of search success.(21)

## **End User Studies**

End user studies differ from many of these studies cited above in that they are often naturalistic studies. These types of studies present difficulties because one cannot control the variables like the level of experience of users or the difficulty of the question. On the other hand these studies provide insight into real user behavior, rather than search performance of experts in controlled settings.

Domenica Barbuto and Elena Cevallos conducted a survey of CD-ROM users at Hofstra Library. Their research showed a marked preference for CD-ROM over print, a high level of user satisfaction, and a poor understanding of alternative resources or the structure of these resources.(22) In another large scale study done at the University of North Carolina, Bucknall and Mangrum surveyed over 1000 users to determine their level of experience, the appropriateness of database choice, their general satisfaction level, preferred source of assistance and patterns of multiple usage.(23) Lancaster also studied end users and compared their search performance to experts. His findings indicate that despite the high level of expressed user satisfaction, the actual search performance of end users as measured by recall is poor.(24)

## **Methodology**

### **Methods for studying user behavior**

There are a large number of sociological research methods which have been invoked in studying users. Some comprehensive studies use multiple methods for collecting data. Kuhlthau, for instance, used user logs, interviews conducted at various points in the search process and she also asked users to draw flowcharts of their research process.(25) The nature of online systems presents additional opportunities for data gathering including online surveys and transaction logs. Numerous studies use transaction logs for obtaining objective recordings of user behavior, and while they provide little insight into the mental decision making process of users, gross patterns often emerge. Tenopir et al.(26) and Sullivan and Seiden(27) used protocol analysis in which one asks users to talk aloud as they perform a task. The former used these studies to confirm the role of affective behavior in information seeking. Coupled with transaction logs and interviews one can gain significant insight into information seeking, but the small number of subjects necessitated by the density of the data can be misleading.

Our study utilized both individual and group interviews. The individual interviews were conducted in the Reference Area in the Library during the Spring and Fall semesters of 1996. Students who were engaged in research using one of the Library's bibliographic databases or Netscape were approached at random and asked if they would consent to be interviewed when they had finished their search session. Each interview took approximately twenty to thirty minutes. The interviewer asked the student to talk about each search session as well their history of computer use and database searching.

A first questionnaire was tested with ten students and subsequently revised to elicit more information about the user's background and used with another thirty two students. The revised questionnaire is included in Appendix A.

The first two group interviews were used to delve further into user's perceptions and understanding of the information environment and the role which prior experience and education plays in developing those perceptions. Eight students participated in each group interview. The research team specifically sought students who had completed at least two major research projects while at Skidmore. The initial sets of questions were similar to many on the questionnaire, but then students were asked to talk about sources they would use to answer 6 different questions. Lastly they were asked to talk about relative navigational difficulties in the electronic and physical library environments.

A third group interview required that participants be in their Senior year and either be engaged in thesis work or have completed a senior thesis during the previous semester. Six students participated in the third interview. In this interview students were asked to complete the revised questionnaire before the start of the interview. It became evident as the data analysis began that a key independent variable in information seeking strategies is type of the research task. The research team wanted to investigate the extent to which this might influence overall information seeking patterns by studying a group of users who were engaged in the same task - a major senior level research thesis. While the findings were interesting, we lack comparable data for other users and this paper will not discuss this data.

The statistical data cited below is drawn from the individual interviews. The focus group findings are used to elucidate and corroborate the data from the individual interviews.

## **Who are the Users?**

Skidmore is a highly selective co-educational liberal arts colleges with approximately 2100 students. The curriculum is strong not only in the traditional liberal arts but also in fine arts and pre-professional programs in business, education and social work. The top five majors are in rank order, Business, Psychology, English, Studio Art and Government. Some disciplines are much more library research focused than others. For example, the business curriculum uses library resources extensively from their first course through their senior seminar, but in other disciplines such as English, students are not typically introduced to research until their senior research paper. It is not unusual to find a senior who does not know how to use the library to do research, but that situation is becoming less common.

The vast majority of our subjects were juniors and seniors (29 of 42), four were first year students and nine were sophomores. This was to be expected because with the notable exception of the business students and some of the basic writing courses few courses require library research early in the major. Students were working on a variety of assignments including senior theses and term papers, company research for the entry level business course, searches for particular documents such as law cases.

Eight students were business majors and another six were government majors, a reflection of both the distribution of majors in the Skidmore population and the patterns of library research required by various majors. Other majors represented were American Studies (4), Psychology (2), English (2), Philosophy (2), Geology (2), Exercise Science (2), Biology (2), Music (1), Education (1), Art (1), History (1), Classics (1), Asian Studies (1), Sociology (1), Economics (1) and two undeclared.

During the interview, students were asked about their computer experience prior to coming to Skidmore. All but nine students had some level of computer experience before entering college. Of these nine students, four had their first experience with computers since coming to Skidmore and for five the data was unclear. In

the focus groups of seniors, several noted that they had been using computers since elementary school. Seventy-five percent of the students used word processing prior to coming to Skidmore, but only 25% used email. Sixteen percent said they used the Web, but only four students said they had prior experience with database searching. Student's pre-college computing experience continues to increase. Skidmore administered a computer literacy survey to all entering first year students in both the Fall 1994 and 1996. Table 1 shows the increase in familiarity and use with word processing, email, and the Web of entering students over this two year period.

Based upon the interviews, we graded students general level of library expertise. This judgment was based upon our impressions of the students' understanding of libraries and their resources. Half the subjects were either beginners or novices, another quarter were intermediates and five were considered to be expert. The data was inconclusive for the remaining four. Yet, with respect to database searching, over 40% had experience searching bibliographic, fulltext and Web databases. Twenty-seven out of the thirty-seven who responded to our question indicated that they had done database searching either several times before or many times before. Only two had never searched before and eight indicated they had searched one time prior to the interview. Database search experience does not appear to correspond with level of library expertise.

When asked how they learned about online searching, we anticipated students would most often state that they learned either on their own or from peers. (Table 2) Earlier studies of end users of both OPACS and Medline had found that most users learned on their own either through trial and error or documentation.(28) However, in our study the most frequently noted means of learning how to search was from a librarian in a one-on-one situation, (nearly 50%), while only 12% noted that they had learned from their peers. Nearly 36% indicated that they had learned on their own. Interestingly, though 50% had formal library instruction, only 17% (7 subjects) indicated that this contributed to their knowledge of how to search. The focus group data showed that among those who relied upon their peers for instruction, the peers played a fairly significant role in teaching searching. Valentine also found that peers were a significant source of guidance among undergraduates.(29) These students estimated that between 40 - 60% of their knowledge about searching came from their peers. But again those who indicated that peers were influential were in the minority.

Many students encounter bibliographic instruction in their courses, but students' exposure to any uniform library experience is unlikely because of the nature of the general education requirements at Skidmore. Those students who are either business or government majors have a strong chance of encountering multiple bibliographic instruction sessions in their career. One of the focus group participants who was a business/government double major said she had three or four sessions and indicated that probably 50% of her knowledge of searching came from those sessions. However, she said that among business majors, the collaborative team approach used in the curriculum allowed for students to become expert in one area and so their knowledge of multiple resources was often limited by the nature of their work.

## **The Information Seeking Process**

### **Database choice**

The digital library at Skidmore is both the collection of commercial bibliographic and journal article databases as well as the heterogeneous, distributed collections of networked information found on the World Wide Web. Students enrolled at Skidmore have access to several hundred bibliographic, factual, and journal article sources including Lexis/Nexis, FirstSearch, and a myriad of individual databases ranging from

Cambridge Scientific Abstracts to Women's Studies Index. Fulltext sources in addition to the Web include Business Periodicals OnDisc, Lexis/Nexis, Project Muse, Ethnic NewsWatch, and Expanded Academic Index.

Users choose databases because someone directed them to it; because of its intellectual content; or because of certain characteristics of the database - such as its format (i.e. fulltext), its currency, its ease of use, or prior familiarity. When asked why they were searching a particular database, most students responded with two or three reasons. We coded only the first two answers. Nearly 40% of the students indicated that they chose a particular source because of its subject area or scope of coverage. Nearly 30% responded that they were directed by a professor to a particular source. (Table 3).

Even among fairly knowledgeable users, ease of use was an overriding concern in database choice. Many stayed away from databases which they perceived as too difficult though there was no unanimity on what databases were considered difficult to use. For instance, some thought that Lexis/Nexis was too difficult to use, while others used it to the exclusivity of even more appropriate sources.

### **Appropriateness of Database Choice**

Allen's study of database choice found a considerable level of inappropriate database use and Bucknall's findings indicated that this is an area for concern.(30) In the Skidmore study, the findings did not confirm that data. Of the forty-two cases, only three database choices could really be said to be inappropriate. The high level of database appropriateness can be explained because of the reliance on librarians, faculty or peers as a means of selecting a resource. However, when asked what other databases they might use, most students did not know, or listed those with which they were familiar whether or not they were appropriate.

The most popular database/service in the study was Lexis/Nexis. It was used by 20 of the subjects. The second most popular service was SilverPlatter which was used by 8. Subjects also reported using the Web (5), FirstSearch (4), and IAC's Expanded Academic Index (3). One student was using Ethnic NewsWatch and one was using ProQuest's Business Periodicals OnDisc. Lexis/Nexis is the library's most heavily used service and the ratio of service use in our samples is corroborated by our monthly search statistics.

That users generally understand the scope of the databases and how they differ by subject coverage was confirmed in each of the general focus groups. Students were asked to state what sources they would use to research 5 different topics covering popular culture, current affairs, and an historical topic. While there was some variety among the answers, and students often displayed creative strategies, none of the sources was inappropriate. For example, when asked where they might go to find out about historical background for Homer's Odyssey, students suggested not only the online catalog, but Project Perseus, as well.

### **The Role of the Web**

Virtually all students (32 out of the 39 who were asked) had experience using the Web and 60% of those noted they use it for research. Most students did not appear to consciously categorize the Web any differently from other bibliographic tools; they considered it just one more resource to add to their bag of tricks. One reason which students consistently pointed to for choosing the Web was the success they have when searching it. One student after unsuccessfully researching "Mexican labor relations" on Business Periodicals OnDisc switched to the Web. He retrieved over 200,000 documents (understandably) because of the relevance ranked/ORed search, whereas the ProQuest software would have searched for his specific terms in that specific order. His reasons for choosing Netscape were because it was the "easiest to use" and

he was most familiar with it. Both the individual interviews and focus groups indicate that most students do not have a clear understanding of what is on the Web. For example, one person noted that all the information on the Web is broad and basic and at the same level as a text book.

### **Search strategies, successes and problems**

The students were asked to describe their search strategy - the terms and/or concepts they used and the search statements entered. In general they recall their searches as fairly straightforward and simple; they made fairly few moves, preferring to enter a search and browse the results with only occasional refinements. They seemed to understand how to combine concepts with simple boolean ANDing, and only five displayed any complex searching involving multiple concepts or proximity operators. Not one of the students used controlled vocabulary to do their searches, a finding which has been borne out by numerous other studies. Students neither use thesauri, nor use retrieved citations as a source for other appropriate terms. Yee concludes that novices don't appreciate a distinction between natural language and controlled vocabulary. (31)

Fulltext searching presents additional challenges. Unfortunately, students do not recognize that most strategies employed in searching bibliographic citation databases are often inappropriate in the fulltext environment. While some students varied search strategies if they were searching a fulltext database (eg. using proximity operators or searching for multiple occurrences of a term), about half of those searching Lexis/Nexis did either single keyword or simple Boolean ANDing. Our findings partly contradict those of Tenopir's and Nahl's study of novices using Magazine Index, ASAP. As Tenopir and Nahl found the most common strategy for searching fulltext was to link using Boolean operators. Yet their study also found that novices used commands that enhance fulltext searching such as proximity searching and field limiting and full-text displays such as KWIC. While we did find students using the last mentioned strategy, differences may be explained by the training their subjects received prior to participating in their research.(32)

Tenopir's and Nahl's work with fulltext databases found that users made moves to limit sets twice as often as moves to enlarge sets. They used KWIC and full-text displays frequently to make decisions about refining searches. There was a tendency to over use narrowing strategies - thus resulting in the retrieval of null sets. (33)

A number of students in our study who retrieved large sets attempted to refine their search. The most frequently used strategies to refine searches were

1. to add additional terms; (ex. One student did the following search in Lexis/Nexis for patterning of group behavior in schools. She entered school and pattern and teacher and student and group and human-behavior and size and class and psychology.)
2. to change the terms they used; or some combination of 1 and 2; (ex. A student who had first tried the search Hawaii and same sex marriages retrieved a very large set; she refined the search by adding homosexual and then refined the search by adding Hawaii once again;
3. to change the database.

Students encountered a variety problems as they searched. Some of these problems were explicitly mentioned by the students, but others were implicit in the data. The most frequent areas where students encountered difficulties were in selection of the appropriate database (48%), utilization of appropriate search techniques (45%), and in selection of appropriate search terms (33%). (Table 4). Knowledge and application of Boolean did not appear to be a problem. Lancaster's data corroborates our findings. He found that the

greatest problem users faced was that they don't identify and use all the terms needed to perform a more complete search, frequently because users search too literally.(34) The tendency to literally translate a topic into a search statement was favored by most users.

Nearly one quarter of the students experienced either technical difficulties or printer problems. Occasionally technical problems were severe enough to impact the search strategy. For example, one student wanted to try Lexis/Nexis and Expanded Academic Index for his topic, but they were unavailable. He ended up trying other icons on the desktop including numerous Silver Platter Databases, but because he was looking for fulltext he did not feel his search was at all successful. However, this was the exception rather than the rule and generally technical problems did not adversely affect search satisfaction.

## **Evaluation of Search Results**

Information retrieval studies identify two major types of feedback - magnitude feedback and relevance feedback which influence users' judgments of results of a search.(35) Tenopir et al. found that searchers emphasize control of set size over relevance. The user first makes a determination about whether or not there is too much or too little information. If the set seems to large or too small, a user may reject the set a priori. At other times, s/he may reject a set that has a small percentage of relevant items based on a cursory review. The optimal set size varies with the purpose of the search, even for the same user.(36)

We inquired as to whether users found too much or too little information and what strategies they used to cope with "magnitude" problems. Two sources in particular, Lexis-Nexis and Netscape frequently retrieved large sets. Students' strategies for coping with too much information differed. Only one student gave up because of too much information. He did a search on latchkey kids and said he could not modify the search to what he considered to be the "right" number of articles. But most students seemed able to cope with information overload remarkably well. The most common strategy was to modify the search by adding terms as noted above. One person who was using PsychInfo said she would have to refocus the search on a particular age group, but she still rated the search as a 5. Another, after switching databases said she found too much in the new database, but that was better than the paucity of references from her earlier search and she rated the search as a 6. A student using Netscape, noted that the issue of set size was not a problem, because one could "ignore" the results.

Transaction logs of students at Carnegie Mellon University using the Proquest system indicated that students generally chose articles from the last 3 years only to print. Since most sets are printed in reverse chronological order, this finding may reflect a strategy wherein students may retrieve substantial sets, but limit the number of citations in the set that they are willing to evaluate.(37)

Though research on relevance shows it to be a complex phenomena involving such elements as topic of the document, subject area, recognition, novelty, quality, orientation, recency, availability and length,(38) the undergraduates in our study based their decisions most often on the citation's topic match to the research question's topic. That subject might be indicated either by the title, abstract, keywords or full text, but students strongly prefer the expediency of determining subject by looking at either the title or headline. Nearly sixty percent used this field alone to determine relevancy. Even among seasoned researchers, the title can be the sole source of a relevance decision if it represents the subject matter well.(39) For those who needed to go beyond the title, they generally used the abstract where available (used by 17%). Less frequently (used by fewer than 3 subjects each) students use keywords, the lead paragraph of a fulltext article, the KWIC format or the entire article. Other criteria for judging relevance included the authority of the

source (12%), date of the material, the language, and the availability of source material in the library. It is likely that many more students use date and language as evaluation criteria than explicitly mentioned it because they are not consciously aware of screening these materials.

Park noted that relevancy is determined by both the attributes of the citation or document and the context in which the evaluation takes place. His work in relevance suggests that three separate types of factors comprise the context: internal factors such as the user's expertise or prior research experience in the problem area, their education, experience and perceptions; external factors which are those particular to the research or search event such as priority of information needs, importance of the information, scarcity of the information, stage of research, and end product of research; and the problem context factors which define how and why the user employs the information to construct or solve the problem.(40) A significant difference between Park's subjects and the typical undergraduate is that the former are acculturated to disciplinary scholarly communication patterns. So they bring their own experience with certain journals, authors or conferences to bear on their relevance decisions. Undergraduates rarely possess this level of understanding of their discipline. In terms of the context factors, external factors seem to play a far greater role than either internal factors or the problem context. However, several students mentioned explicitly that their own disciplinary knowledge played a major role in determining relevance. Others noted that they relied upon external cues from faculty who have taught them about the key journals in the discipline and where the "good work" is published.

Students were generally pleased with their searches. Twenty-eight rated their searches 4 or higher on a scale of one to six when asked to rate their searches from poor to excellent on a six point scale. (See Table 3) This high level of satisfaction has been found in most other CD-ROM end user studies even when the results of the search weren't particularly useful. Lancaster calls this the "false confidence syndrome."(41)

Among users who were unsuccessful, their perceptions as to why a search failed rarely had anything to do with search strategies or search terms, although these were frequently to blame. Rather satisfaction is a complex phenomena which is dependent upon the user and their interaction with the system. Louise Su suggests that a user's judgment of system success is a multidimensional process incorporating as many as twenty six different aspects. Only four of these factors directly relate to relevance. Other aspects which can influence satisfaction are search efficiency, system capabilities and coverage, accessibility, and output.(42)

These factors are evident in the reasons that the students gave for why their searches were less than successful. Of the ten students who rated their search three or lower, five said there was too little information, two indicated that they were dissatisfied because the articles were not in the library. One other factor Su mentions as influencing satisfaction is the extent to which a search meets a user's expectations in terms of the number and type of materials retrieved.(43) We found that our users generally had very accurate expectations concerning the type and format of materials that would be contained in the database they were searching, but among those who rated searches 3 or lower, three students had certain expectations with respect to the type of materials to be retrieved which went unmet. Each of these was hoping to find the complete articles online and were disappointed to only get citations.

That students were consciously or unconsciously operating with these dimensions in mind was apparent from their responses to the question as to why they chose to do research using computer based resources. Like other studies,(44) students most often cited convenience (38%) and efficiency (43%). Six students (14%) mentioned that they liked to use computers because computers allowed for one-stop-shopping. This answer was not coded as convenience because it often brought in the parameter of fulltext. Seven (17%) students did specifically mention that fulltext was a key reason for using computer databases. Others cited currency of the data (14%), completeness of the data (12%), that online databases permitted users to

connect concepts in ways that were not available in print resources (i.e. boolean capabilities), (7%), their familiarity with the use of online databases as opposed to print materials (5%), that they were required to use online databases (5%), that online databases were readily available (2%) and that one could print from online databases (2%). Bucknall and Mangrum found that only 18% of users cited features related to content as a reason for choosing to search CD-ROMs.

## Discussion

Information seeking is a highly complex task involving the interaction among the user, the information need, and the information resources. The heterogeneous environment which undergraduate users confront in academic libraries today increases the complexity, as the not only the format of information, but the number of resources seems to grow exponentially. In the traditional library, undergraduate students dealt with a fairly limited spectrum from the range of information sources - some basic reference sources, books, magazines, a few major newspapers and scholarly journals. The current digital environment expands the sources that students encounter in their research process and expands the formats in which they will encounter information requiring the student to navigate back and forth from the digital to the print/microform environment. For example, a Geology major mentioned that he used a Geology newsgroup to look for information on a topic and another American Studies major said that she has regularly been using a newsgroup on Vietnam veterans for her research in this area.

Undergraduate students are at a certain disadvantage in coping with this environment, because they have not yet been acculturated to the scholarly communication patterns of a particular discipline. Furthermore, they are often required to do research in very disparate subject areas, so even if they are beginning to develop an understanding of disciplinary norms in their major, they usually do not possess that level of understanding across all the areas in which they may be doing research. Our data seems to indicate that while students seem to be able to get by, they still lack knowledge critical to becoming effective researchers.

One indication of the relatively limited understanding of this environment is the level of granularity at which most users perceive this environment. As indicated in their strategies for choosing a particular database, they perceive highly complex data services like Lexis/Nexis or FirstSearch as a single resource and often are not cognizant that they have actually chosen to search particular databases within these environments. In the same way, they see the Web as one large information space. We have observed that users distinguish among different sources if they appear as distinct icons on the desktop.

Students often begin their research process at the computer. Amanda Spink found that nearly 80% of users of CD-ROM and OPACs were at the beginning of their search.(45) Valentine found that college students tended to start their research with something familiar(46) and among our students the sources with which they are most familiar are digital in nature. Since we didn't interview students who were using the OPAC, we didn't have data directly relating to the role of the OPAC in students' research. The OPAC appeared consistently on the list of "other" databases which students used, but students in the individual interviews tended not to mention it unless specifically asked. When we did ask focus group participants if they used the OPAC, those students were virtually unanimous in it being a standard first step in most research problems - so obvious that they didn't think about it.

The large majority of students possess a high level of comfort with computers and so do not see the technology as a barrier to research. Rather, they indicate a strong preference for computer based resources in as much as they enhance efficiency - the one-stop shopping. Consistent with Valentine's assessment,(47) the data from the study indicates that novices are driven largely by expediency and convenience in

developing their information seeking strategy. For example, because of the importance of one stop shopping, convenience and efficiency many students turn to fulltext sources like Expanded Academic Index, Lexis/Nexis, and the Web when other sources would be more appropriate.

The focus group data shows that even among more sophisticated library users these are strongly motivating factors. Valentine also found the same strong motivation among her focus group participants. She noted that students chose research methods which would get them in and out of the library as quickly as possible. She also wrote that some of the students she interviewed perceived electronic sources to be the best option for obtaining fast information.(48) In fact, students in our focus group noted that the physical library presented more barriers to their research process than the digital library. They mentioned the difficulty in finding what they needed "on the shelf," the problems with photocopiers, and that the LC classes were not in the expected order, as problems encountered in the physical library. One focus group participant noted that his recovery strategy from navigational problems in the physical library was to utilize digital sources.

Overall there is a strong preference for digital sources. This preference is reinforced by a lack of familiarity with print sources beyond such secondary school tools as Reader's Guide to Periodical Literature. Bucknall and Mangrum found that 26% of his survey respondents said they would never use printed indexes or abstracts and another 27% they would only use those sources if they were the only thing available.(49) The more familiar the user is with a system, the more they will grow to rely upon it. In Bucknall's and Mangrum's study, he found that the most experienced users were driven to use sources not because of content, but because of the medium, itself.(50)

Students seem quite unaware of the complexity of the digital environment and they tended not to realize the exact nature of the demands that the environment places upon the user in order to be successfully exploited. This oversimplification may be due in part to the physical environment in which the users find themselves. Ironically the environment implicitly contradicts the "heterogeneity" characteristic because to the end user, the desktop computer presents "world of information" as a unified/homogenous world. And this is becoming ever more so as larger numbers of databases are available through a common interface or browser. This singular mode of access may make it more difficult for the user to discern substantive differences among the sources available from the desktop. It may blur not only the distinctions among different types of sources such as bibliographic, full-text and hypermedia, but it may also blur differences among the intellectual characteristics such as the sources' authority, depth and scope of coverage.

Our research and that of others cited throughout this paper provide strong evidence that even a single online or CD-ROM system requires an understanding of numerous factors. In Solomon's research on expertise, he notes three areas of knowledge which define expertise with a single system: 1) capacities, limitations and range of use of a system; 2) procedures and steps required to use the system; and 3) strategies that make the system achieve its objective.(51)

Nearly two-thirds of the subjects had used citation, fulltext and Web based sources, so we can assume at minimum that they have used three different data services and several students listed as many as six different services. We also know from their strategies that few possess an expert's understanding of even one of these services. Students learn the minimal amount about a system to enable them to interact with the system and retrieve something. Because this knowledge usually entails a basic keyword and boolean strategy it can be utilized in virtually all the search systems they are likely to encounter with some success.

The digital library environment does appear to change help seeking patterns. In fact, the data shows that nearly 50% of the students admitted seeking help at some point in the search. Eighteen of 39 responding to this question said they sought help from a librarian. While Valentine noted that students in her study relied

heavily upon peers, only 1 student cited a peer as a source of help, and one person cited a faculty member. Valentine noted that among those who did ask for help from instructors or librarians, upper division undergraduates tended to ask their instructors while lower division students turned to librarians for assistance,(52) but our data didn't reveal any difference among students.

Kuhlthau found that students relied on formal mediators such as librarians or teachers primarily to assist in locating sources in the library's collection rather than on topic formulation.(53) However, we found that half the students asked for help at search initiation. Since our students frequently turn to librarians for assistance, these interactions do provide an opportunity for the librarian to work with the student on topic refinement.

## **Models**

Most information retrieval research has grouped subjects into one of three categories: general end users; end users with subject knowledge; and expert or professional searchers. We find that this model which places the general undergraduate in the first category is too simplistic. The undergraduate does not present a singular behavior although some behaviors are more common than others. The individual interviews and focus groups do indicate at least two distinct levels of users: the novice and the sophisticated and one can hypothesize that there are others who fall somewhere between these two stages. These distinctions do not appear to be predicated strictly on experience. In fact, if one looks at two users who have had the same number of research assignments over the course of their career, the two sets of behavior patterns are still evident.

Those behaviors are:

- Novices' understanding of a database is usually limited to the subject of the database; they are generally unaware of the years covered in the database or the differences among databases as pertains to types of sources covered.
- Novices often choose a particular database because someone (librarian, faculty, peer) told them to use that database; when asked what other databases they might use for the search, they either don't know any other database or list others with which they might be familiar whether they are appropriate or not.
- Novices rely upon strategies which have worked before; they are generally aware of only one disciplinary resource, but may also try some multidisciplinary sources such as IAC or the Web if the disciplinary source proves unsuccessful.
- Novices search strategies are fairly simple, either using keyword, phrase or two or three words with an AND connector; they tend not to vary a strategy as they move from database to database. They can learn and apply higher level search strategies, but this expertise in a database is procedural, not conceptual.
- Novices tend to be literal in their search statements, rarely using synonyms. For example, a student looking for early anabaptist thought used the terms anabaptist, early thought, Europe.
- If the user finds too much information, he/she will either browse the results without refinement or redo the search with additional key terms. If they don't find enough information they either change databases (or remove some terms).
- Relevance judgments are made based upon the subject of the question and utilizing the title field. Again in assessing relevance, the novice user typically uses the title to determine if the material is

on target. Thus, they do little evaluation online and anecdotal data suggests that many simply photocopy the materials or make interlibrary loan requests without further evaluation. Efficient evaluation seems to be motivating factor.

- The novice's model of the information environment is overly simplified. They do not possess a knowledge of how different sources of information are structured.

An example of novice behavior was exhibited by a senior philosophy major doing research on violence in Buddhist countries. She had chosen to use the Web for her research, because she said she always found something. When questioned about her use of other sources, in particular her use of Philosopher's Index, she said she never heard of it, nor would she expect philosophers to have an index, because philosophers don't even have jobs.

- Sophisticated users are able to vary the search strategy with the database and can often exploit features of the database.
- They are aware of the range of resources in the Library and what's not in the Library.
- They typically use library collections broadly and browsing behavior is a significant means of finding information. They are comfortable trying various strategies in order to find the best material.
- They don't hesitate to ask librarian for assistance as they move from database to database or to develop more complex search strategies.
- They may not know how specific sources are structured but understand the overall information environment.
- They take into consideration multiple aspects of the source when evaluating it - eg. the journal in which it appears, the author, even the country of origin. Relevance is also based upon what they know about the topic
- The sophisticated user tends to possess a more accurate model of the information environment. They are aware of the complexity of database structures, though they may not know the specific procedures for using a particular source.

An example of a particularly sophisticated strategy was given by an art history major who noted that she tended to go to books and use their bibliographies to find seminal articles on the topic. She would then search for that article in either RILA or Art Index and look at the descriptors assigned to the article and search on those terms. Another student used psychology texts for background research and to identify specific researchers and then searched for those researchers as authors.

The novice user may or may not ever "mature" beyond their novitiate stage. The findings about the process of the development of expertise in the use of information systems indicate that this process is not just a matter of acquiring more factual knowledge, but involves a transition or reorganization of knowledge representation.(54)

If not from experience, from whence do these differences arise? Meadow provides a schema that may be more helpful in distinguishing differences among undergraduate end-users - one based upon knowledge and skills which impact directly upon search performance. Because we are studying a more complex phenomenon than the user's interaction with a single IR system, we can hypothesize that there will be many more types of knowledge and skills which distinguish between the novice and expert than in Meadow's

model.(55) In fact, it is likely that expertise may exist on different levels. Thus a novice may be procedurally proficient in searching a single database, but not possess an accurate model of the broader information environment and so on another level, they remain a novice.

## **The Role of the Faculty**

Our research does provide some clues as to how that maturing may occur. A junior psychology major described personal searching strategies that are more typical of faculty - looking for known authors with expertise in particular areas; browsing journals in a specific subfield. Faculty mentoring is possibly one way the student begins to incorporate more of these sophisticated strategies.

When asked if any faculty encouraged or required computer based searching subjects mentioned a wide variety of courses and faculty. Faculty in the government and business departments were cited most often, but students also noted faculty or courses in American studies, classics, philosophy, math, biology, social work, English, psychology, exercise science and German. For the past three summers, Skidmore has sponsored a ten day workshop for faculty on integrating technology into teaching and research. A key goal of this workshop has been to raise awareness of changing patterns of scholarly communication within different disciplines. The overlap between those faculty whose names are cited by students and those enrolled in the workshop is nearly 100%, an indication that the workshops have at least been successful in raising awareness of the library's digital resources.

But the faculty's role goes well beyond simply pointing out appropriate resources. While Kuhlthau's research findings indicated a minimal role for formal mediators - librarians and teachers - Fister's research on successful researchers and ours indicates that faculty at undergraduate institutions do play a significant role in helping students' focus topics. Fister also found that her researchers more often followed faculty information seeking patterns.(56) While subject expertise does not appear to have a significant impact on the search process, acculturation into a discipline does change information seeking behavior broadly. Once students fully understand how information is created, disseminated and organized in a particular discipline, they may possess a better mental model on which to base their heuristics for information gathering in other disciplines and in helping them develop an understanding of some disciplinary norms. Finally, students rely upon faculty to help them evaluate resources(57) and faculty feel keenly that it is their responsibility to help their students develop filtering and evaluation strategies.(58)

## **Conclusion**

The "digital library" appears to have a significant impact on information seeking behavior of the majority of library users.

- Students driven by time pressures and seduced by the convenience of one stop shopping increasingly rely on fulltext and the Web to the exclusion of other resources; Students frequently start and end their research at the computer.
- Even those who use citation sources often focus their entire literature search on computer based resources thereby missing other valuable sources that either fall outside the scope or timeframe of computer based sources.
- Students are have increasing difficulty discerning critical differences among various sources of information because they are available via a single computer and a uniform interface or browser
- The "digital library" does not remedy a poor understanding of the overall information environment. Rather, the heterogeneity of the environment exaggerates and magnifies these problems.

- On the positive side, the availability of "unpublished information" on the Web and virtual "invisible colleges" via newsgroups and listservs allows students to have access to information sources previously only tapped by faculty.
- Because students seek help from librarians both at the beginning and in the middle of their searches, the help seeking provides an opportunity for education about not only the system, but information seeking strategies in general.

These findings have implications for our instructional programs. As technology has gained an ever large foothold in the domain of disciplinary research, faculty have turned to librarians to provide themselves and their students with pragmatic, procedurally focused instruction. Basically they have asked for us to teach only how to search the databases. Technology, as Cerise Oberman has pointed out, has led librarians to forsake teaching concepts in favor of procedure.(59) Yet as our research has demonstrated, while undergraduates can certainly grasp the complexities of databases like Lexis/Nexis, most do not have an understanding of the information environment - an accurate mental model - that will allow them to develop heuristics which they can apply as they navigate through the environment.

The developing digital library demands users have a sophisticated model of the information environment in order to be able to appropriately utilize the wide variety of resources available and to move among search systems which might have not only different interfaces, but distinctly different structures and content. Our findings show that students cannot gain this knowledge through experience alone. It must be taught and to do that we will need to coopt the faculty, to work with librarians on developing assignments which will provide students with a deeper understanding of the information environment rather than simply providing a knowledge of how to use specific tools.

## **Future Research**

This research study raises more questions than it answers. What it does describe are some general characteristics of undergraduate searchers and patterns of information use. Because we have only taken a snapshot of one moment in time in the research process and because we are dependent upon the student's recall of what he/she did we do not have an accurate picture either of the undergraduate's search process over time, or the way in which these sources are integrated into their overall information seeking behavior. Furthermore, we can't say definitively whether our results are applicable outside of Skidmore because the number of subjects interviewed was so small.

We need to do both more narrowly focused studies on such issues as the contextual factors which influence relevance decisions of undergraduates, and broader studies which look at the entire information seeking task and the nature of the use of digital resources for different types of assignments and different disciplines.

Secondly, since our research recognizes at least two levels of users and different mental models of the information environment, we need to develop a better understanding of the sophisticated user and the factors which distinguish them from so many of their peers.

## **NOTES**

1. Ann Bishop, "Building a University Digital Library: The Need for a User-Centered Approach." (Urbana-Champaign: University of Illinois (1995). Online. Available: [http://anshar.grainger.uiuc.edu/dlisoc/monterey.final\\_copy.html](http://anshar.grainger.uiuc.edu/dlisoc/monterey.final_copy.html) ([http://anshar.grainger.uiuc.edu/dlisoc/monterey.final\\_copy.html](http://anshar.grainger.uiuc.edu/dlisoc/monterey.final_copy.html)).

2. Nancy Van House, "User Needs Assessment and Evaluation for the Berkeley Electronic Environmental Library Project" in *Proceedings of Digital Libraries: The Second Annual Conference on the Theory and Practice of Digital Libraries* (College Station, TX: Texas A&M University 1995) 71-76; Online, Available: <http://csdl.tamu.edu/DL95/papers/vanhouse/vanhouse.html> (<http://csdl.tamu.edu/DL95/papers/vanhouse/vanhouse.html>).
3. Ann Bishop, "Artists on the Internet," in *INET '95 Conference Proceedings*, vol. II (Reston, VA: Internet Society 1995) 1009-1018.
4. Peggy Seiden. "User Expectations and the Digital Library" (paper presented at the Spring Meeting of the New England ACRL, Worcester, MA, March 15, 1996) 10.
5. Carol Kuhlthau, *Seeking Meaning: A Process Approach to Library and Information Services* (Norwood, N.J.: Ablex Publishing Company, 1993) 37-53.
6. Barbara Fister, "The Research Processes of Undergraduate Students," *The Journal of Academic Librarianship* 18, no. 3 (1992) 163-167.
7. Barbara Valentine, "Undergraduate Research Behavior: Using Focus Groups to Generate Theory," *The Journal of Academic Librarianship*, 18, no. 5 (1993) 300-304.
8. Fister, 164., Kuhlthau, 46, 52-53.
9. Fister, 167. Mellon, Constance, "Process not Product in Course-Integrated Instruction: A Generic Model of Library Research." *College & Research Libraries* 45 (1984) 471-478.
10. Riya Fidel and Dagobert Soergel, "Factors Affecting Online Bibliographic Retrieval: A Conceptual Framework for Research," *Journal of the American Society for Information Science*, 34, no. 3 (1983) 163-180.
11. Tefko Saracevic et al., "A Study of Information Seeking and Retrieving: I. Background and Methodology," *Journal of the American Society for Information Science*, 39 (1988) 161-175.
12. Ingrid Hsieh Yee, "Effects of Search Experience and Subject Knowledge on the Search Tactics of Novice and Experienced Searchers," *Journal of the American Society for Information Science*, 44, no. 3 (1993) 161-174.
13. Charles T. Meadow, Jiabin Wang and Weijing Yuan, "A Study of User Performance and Attitudes with Information Retrieval Interfaces," *Journal of the American Society for Information Science*, 46, no. 7 (1995) 504.
14. Carol Tenopir, Diane Nahl-Jakobovits, and Dara Lee Howard, "Strategies and Assessments Online: Novices' Experience," *Library and Information Science Research*, 13 (1991) 237-266;
15. MaryEllen Sievert and Emma Jean McKinin, "Why Full-text Misses Some Relevant Documents: An Analysis of Documents Not Retrieved by CCML or Medis," in *Proceedings of the 52nd ASIS Annual Meeting* (Medford, N.J.: Learned Information, Inc. 1990), 34-39.
16. Barbara Wildemuth et al., "Search Moves Made by Novice End Users" in *Proceedings of the 55th Annual ASIS Meeting*, (Medford, N.J.: Learned Information, Inc. 1992) 154-160.
17. Blaise Cronin and Carol A. Hert. "Scholarly Foraging and Network Discovery Tools." *Journal of Documentation*, 51, no. 4 (December 1995), 390-391.
18. Spink, Amanda, "Interaction with Information Retrieval Systems: Reflections on Feedback" in *Proceedings of the 56th ASIS Annual Meeting* (Medford, N.J.: Learned Information, Inc. 1993) 117-

120.

19. Taemin Kim Park, "The Nature of Relevance in Information Retrieval: An Empirical Study," *The Library Quality* 63, no. 3 (July 1993) 318-351,
20. Kuhlthau, 38.
21. Louise Su, "Is relevance an Adequate Criterion for Retrieval System Evaluation: An Empirical Inquiry into User's Evaluation," in *Proceedings of the 56th ASIS Annual Meeting* (Medford, N.J.: Learned Information, Inc. 1993) 95-98.
22. Barbuto, Domenica and Elena Cevallos. "End User Searching: Program Review and Future Prospects," *RQ*, 31, no. 2, (Winter 1991) 214-227.
23. Tim Bucknall and Rikki Mangrum, "U-Search: A user Study of the CD-ROM Service at the University of North Carolina at Chapel Hill," *RQ*, 31, no. 4 (Summer 1992), 542-553.
24. F. W. Lancaster et al., "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, no.3 (Spring 1994) 377-383.
25. Kuhlthau, 81-90.
26. Carol Tenopir, Diane Nahl-Jakobovits, and Dara Lee Howard, "Magazines Online: Users and Uses of Full Text" in *Proceedings of the 52nd ASIS Annual Meeting* (Medford, N.J.: Learned Information, Inc., 1989) 173-175.
27. Patricia Sullivan and Peggy Seiden, "Educating Online Catalog Users: Assessment of Needs," *Library Hi Tech*, 3, no. 2 (1985) 11-19.
28. Charles T. Meadow, Gary Marchionini and Joan M. Cherry, "Speculations on the Measurement and Use of User Characteristics in Information Retrieval Experimentation" *The Canadian Journal of Information and Library Science* 19, no. 4 (December 1994) 5-7.
29. Valentine, 303.
30. Gillian Allen, "Research Notes: Database Selection by Patrons Using CD-ROM" *College and Research Libraries*, (January 1990) 69-75; Bucknall, 544.
31. Yee 169.
32. Diane Nahl and Carol Tenopir, "Affective and Cognitive Searching Behavior of Novice End-Users of a Full-Text Database" *Journal of the American Society for Information Science* 47, no. 4 (April 1996) 285; Tenopir et al. 173.
33. Nahl, 281.
34. Lancaster 379-380.
35. Spink 115.
36. Tenopir, "Strategies and Assessments Online" 244, 252.
37. Troll, Denise.
38. Peiling Wang and Dagobert Soergel, "Beyond Topical Relevance: Documentation Selection Behavior of Real Users of IR Systems" in *Proceedings of the 56th ASIS Annual Meeting* (Medford, N.J.: Learned Information, Inc. 1993) 87-88.
39. Park 330.
40. Ibid. 333-341.

41. Lancaster 382.
42. Su 96-97.
43. Ibid. 96.
44. Bucknall 546.
45. Amanda Spink, "Multiple Search Sessions Model of End-User Behavior: An Exploratory Study"  
*Journal of the American Society for Information Science* 47 no. 8 (August 1996) 606-607.
46. Valentine 302.
47. Ibid..
48. Ibid..
49. Bucknall 548.
50. Ibid..
51. Paul Solomon, "On the Dynamics of Information System Use: From Novice to?" p. 163.
52. Valentine 303-304.
53. Kuhlthau 41.
54. Solomon 163.
55. Charles T. Meadow, "Speculations on the Measurement and Use of User Characteristics" 2-7.
56. Fister 168.
57. Fister 166.
58. Seiden 16.
59. Cerise Oberman, "Library Instruction: Concepts and Pedagogy in the Electronic Environment," *RQ*  
35, no.3 (Spring 1996) 320.