

A Study of Factors That Affect the Information-Seeking Behavior of Academic Scientists

Xi Niu and Bradley M. Hemminger

School of Information and Library Science, University of North Carolina, Chapel Hill, NC 27599–3360.

E-mail: xiniu@email.unc.edu; bmh@ils.unc.edu

In an effort to understand how academic scientists seek information relevant to their research in today's environment of ubiquitous electronic access, a correlation framework is built and regression analysis is applied to the survey results from 2,063 academic researchers in natural science, engineering, and medical science at five research universities in the United States. Previous work has reported descriptive statistics about these scientists' information-seeking behavior. This study extends that work to examine relationships between scientists' information-seeking behaviors and their personal and environmental factors. Several regression models, including the Poisson model, the logit model, and the ordered logit model, are built to interpret the correlation among scientists' behaviors. In addition, exploratory factor analysis is used for data reduction. Overall, many factors were found to affect the specific information-seeking behaviors of scientists, including demographic, psychological, role-related, and environmental factors. Of the factors having an effect, academic position was the most important determinant of information behavior.

Introduction

The arrival of various electronic resources, including e-journals, online databases, and digital libraries, has significantly changed the way scholarly communication takes place. Nowadays, academic researchers have convenient and easy access to an unprecedented amount of electronic scholarly resources while also facing a dizzying array of different types of content delivery, including books, journals, web pages, and online databases. In an effort to understand how scientists are adapting to the increased use of electronic scholarly communication and what effect this has on their information-seeking behaviors, the NeoRef research group (<http://neoref.ils.unc.edu/>) at the University of North Carolina at Chapel Hill has conducted a national survey of 2,063 academic scientists from five research universities

in the United States. We have previously reported on this study, the survey methodology, and descriptive statistics of scientists' information-seeking behavior across five institutions (Hemminger, Lu, Vaughan, & Adams, 2007; Niu, Hemminger, & Lown, 2010).

The focus of this article is to understand, for academic scientists engaged in scholarly information searching, the interactions between attributes of the user, their environmental context, and their information-searching behavior. An exploratory analysis using regression techniques was performed to identify the correlations among information-seeking behaviors. Poisson regression, logit regression, and ordered logit regression analyses are applied to identify the specific linkages between contextual factors and information-searching and use behaviors. A framework is proposed for understanding the interactions. The goal is to discover correlations between factors and behaviors and to be able to predict specific behaviors when subject and environmental context are known.

Related Work

Background work related to information-seeking behavior for academics and the survey methodology are described in a previous publication detailing the national study (Niu et al., 2010). This section describes prior research results and methods that have examined correlating factors for explanations of information-seeking behaviors, which are relevant to the analyses performed in this article. In addition, references also are made in the results section to draw comparisons between the results of this study and prior experiments.

A significant amount of research during the past 10 years has focused on how academics have responded to the transition from print to electronic scholarly communications. This research has shown a trend in which researchers are becoming more accepting of electronic resources and increasingly utilizing them as their primary sources (e.g., see the excellent studies involving multiple universities spanning large time periods by Tenopir and King, including King, Tenopir,

Received March 16, 2011; revised August 23, 2011; accepted August 24, 2011

© 2011 ASIS&T • Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.21669

Montgomery, & Aerni, 2003; Tenopir & King, 2002; Tenopir, King, & Bush, 2004, as well as others, e.g., Dillon & Hahn, 2002).

Demographic Effects

Most of the studies mentioned previously have used descriptive statistics to provide a “fact book” of scientists’ information behavior to determine interactions between information behaviors and demographic factors. Howard, Rainie, and Jones (2002) found that people with different sociodemographic characteristics look for different content online. Their review of survey research on Internet use found “significant differences in use between men and women, young and old, those with different race and ethnicity, and those of different socio-economic status (p. 45)”. Other researchers have examined the effect of specific demographic attributes on information-seeking behavior. For example, gender difference was investigated by Boneva and Kraut (2002), Brynin (2006), Kennedy, Wellman, and Klement (2003), and Rainie, Fox, Horrigan, Lenhart, and Spooner (2000). They found that women tend to use the Internet more for social activities and less as an information resource. Losh (2004) identified digital gaps of access and use relating to gender, educational level, and labor force participation. In terms of a generational gap, Raban and Brynin (2006) concluded that older adults tend to lack key cognitive resources, reducing their willingness to utilize information technologies. Hargittai (2002) addressed an educational difference for accessing varying content on the web. In particular, he found that college-educated Internet users consistently use the web for work, education, and other informational tasks, and they were less likely to use the Internet for entertainment purposes. Income and racial differences are more likely explained by other demographic factors such as education (Mullis, Mullis, & Cornille, 2007).

Multivariate Analyses

A series of studies by Palmer (1991a, 1991b) employed multivariate techniques to explore the influence of personality, discipline, and organizational structure on the information behavior of biochemists, entomologists, and statisticians. Most of this research has dealt with everyday information seeking, which could be different from academic information seeking for scholars. A limitation of these studies is that they addressed only one aspect, such as gender, and did not address interactions between aspects. Buente and Robbin (2008), however, simultaneously focused on multiple factors and then applied logistic regression techniques to demonstrate the relative effect of different demographics and information behaviors on whether a user engaged in a particular Internet activity, (e.g., getting online news). Their research has suggested that attributes of the searcher play a role in addition to context. A limitation of their work was the choice of a binary model, which simplified the real situation by treating each specific information act as either “doing” or “not doing.”

This study analyzes the comprehensive data from our national survey to investigate the interaction between users, their environmental context, and their information-searching tasks. This is done by including regression analysis and factor analysis to examine potential factors in predicting a specific information act. Exploratory factor analysis is used for data reduction for some questions.

Framework

Based on Wilson’s (1997) revised general model of information behavior and Buente and Robbin’s model (2008) of everyday life information practices, we propose a simplified model capturing context for predicting scientists’ information behavior. In the proposed framework shown in Figure 1, a person’s specific information action is determined by a combination of attributes of the person, his or her environmental context, the information need, and his or her general information behavior excluding the current specific action of interest. Psychological, social-role-related, and demographic characteristics are treated as the attributes of the person (Wilson, 1997). In this framework, each *specific information action* is treated as a dependent variable in the regression analysis, and the other components are treated as the predictors. This allows each individual information-seeking behavior to be evaluated for correlations to all the other external and internal factors shown in Figure 1. The dashed line signifying a relationship between specific information action and information behavior in Figure 1 indicates that the specific information action is “a part of, but is excluded from” their general information behavior. Excluding environmental and personal factors, searching for scholarly information is assumed to be generally similar for academic scientists in university settings. For this survey, information behavior was further subdivided into information searching, information using, and information collecting. The key variables identified by each component of the framework are shown in Table 1. While some of the predicting variables are correlated, thus reducing the overall predictive power of the analysis, each of the included variables is separately analyzed because it has its own independent effects. For instance, age and position are positively correlated, but neither fully predicts the other.

Methodology

Data Collection

A series of surveys on information-seeking behavior was conducted during the past 4 years to collect data from five research universities in the United States: the University of North Carolina at Chapel Hill (UNC), the University of Florida (FL), the University of Oklahoma (OU), Colorado State University (CSU), and the University of South Florida (USF). The survey design, methodology, and data collection are detailed in a description of the national study (Niu et al., 2010) and briefly summarized here. The survey consists of six parts: Background Information, Types of Resources Used,

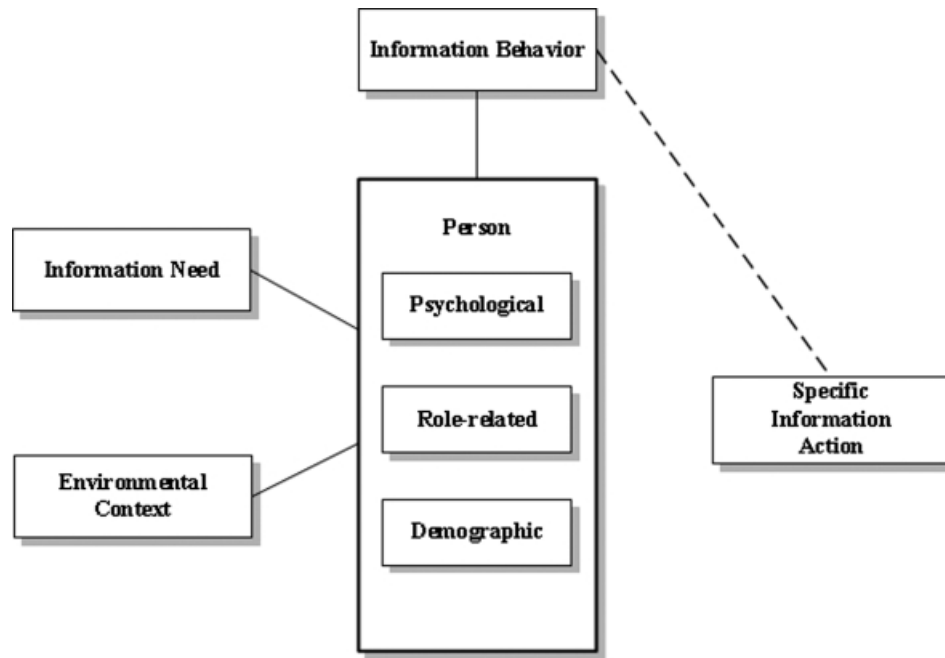


FIG. 1. The proposed framework.

TABLE 1. Description of key variables.

Variable	Description
Demographic	
age	Age of the participant
gender	Gender of the participant
Psychological	
confidence	How confident the participant is that he/she is finding everything
Role-related	
position	The academic position the participant holds (e.g., Assistant professor, PhD student)
Environmental context	
department_type	The department type in which the participant is situated (e.g., natural science, engineering)
distance_to_library	How far is it from the participant's office to the campus library he/she uses most often
Information-searching Behavior	
preferred_search_format	How the participant would prefer to search for information (print or electronically)
number_of_library_visit	How many times the participant visited the library in the last 12 months
search_page	Which interface (Google or the library's home page) the participant would rather use to begin his/her search process
Information-using behavior	
reading_hours_per_week_per_week	In a typical week, how many hours the participant spends reading information relevant to his/her work
percentage_of_articles_with_notes	Of articles in the participant's article collection, to what percentage have notes been added?
preferred_reading_format	How the participant would prefer to read information (print or electronically)
Information-collecting behavior	
maintain_article_collection	Whether the participant maintains a personal article collection
maintain_bibliographic_database	Whether the participant maintains a personal bibliographic database
percentage_of_articles_in_bibliographic_database	What percentage of the articles from the participant's collection are in his/her bibliographic database

Keeping Current, Searching for Information, Personal Article Collection, and Searching & Using Information. The survey questions attempted to quantify the transition of academic scientists to electronic forms of communication, and how this has affected different aspects of information seeking. A copy of the final multi-institution survey is available at <http://ils.unc.edu/bmh/isb/National-ISB-Survey.pdf>

In total, 2,063 academic scientists were sampled from UNC, FL, OU, CSU, and USF, including faculty, research staff, and students (graduate and postdoctoral) in departments including natural science, engineering, and medical science. Table 2 shows the sample size and the response rate at each site. Response rates were high as compared to similar web-based instruments, and the sample population, while not

TABLE 2. Sample size/response rate.

	UNC	FL	OU	CSU	USF
Participants/recruited	902/3,523	423/3,777	139/589	258/4,521	315/6,257
Response rate	26%	11%	24%	6%	5%

UNC = University of North Carolina at Chapel Hill; FL = University of Florida; OU = University of Oklahoma; CSU = Colorado State University; USF = University of South Florida.

TABLE 3. Department-type distribution.

	UNC	FL	OU	CSU	USF
Natural science	36%	50%	100%	66%	46%
Engineering	1%	15%	0%	14%	14%
Medical science	63%	35%	0%	20%	40%

UNC = University of North Carolina at Chapel Hill; FL = University of Florida; OU = University of Oklahoma; CSU = Colorado State University; USF = University of South Florida.

randomized, was representative of the complete populations across the demographic variables (Niu et al., 2010).

Approximately 40% of the respondents were between the ages of 20 and 30 (mostly graduate students), and the rest were fairly evenly distributed by decade. Gender distribution is balanced in the samples from UNC, CSU, and USF whereas there were approximately 60% males from FL and OU. Regarding academic position, UNC, UF, and OK respondents included a large percentage of doctoral-student participants and a small fraction of master's students. Responses from CSU and USF included higher percentages of master's students, and CSU included significantly more research/adjunct staff than did the other schools.

The survey recruitment was done by departments within each university. More than 100 departments across five universities were surveyed. Almost all of the responses were from the targeted departments in natural science, medical science, and engineering. Responses from departments outside the study scope (social science and humanities) were excluded from the analysis. After comparing and combining each department across institutions, 50 majors were identified and divided into the three types: natural science, medical science, and engineering. Distribution of department types for each school is shown in Table 3. The breakdown of participants by individual department is available in the earlier article (Niu et al., 2010).

Data Analysis

The purpose of this article is to examine the effect of user attributes and environmental context factors on the information-seeking behavior of scientists. Discovering correlations between factors and behaviors could lead to the ability to predict behaviors given user attributes and environments using the proposed framework. For example, it would be useful to know whether academic scientists who prefer to use Google as a search engine are more likely to read online

rather than read print materials. Similarly, do researchers who maintain a bibliographic database read more than those who do not? Attempts to understand the relationship between information behavior and contextual factors rely on models to reveal the relationships between a dependent variable (specific information action) and a set of independent variables (contextual factors and behavioral characteristics). Regression analysis is applied to reveal these interactions. Since all the data except the last three open-ended questions from the survey are either numerical or categorical, generalized linear regression models (GLMs) have been employed for analysis. The GLM is a flexible generalization of linear regression by allowing the linear model to be related to different types of outcomes (dependent variables) via a link function, such as a log transformation. The GLM provides a way to unify various statistical models, including linear regression, logit regression, and Poisson regression, under one framework (Nelder & Wedderburn, 1972). Specifically, Poisson regression is used for a numerical variable as the dependent variable, logit regression for binary dependent variables, and ordered logit regression for multilevel categorical data.

One challenge of working with a large series of data is synthesizing the findings. Exploratory factor analysis is applied to survey questions which have a series of subquestions for reducing the large set of variables to a smaller set of exploratory factors. Factor analysis, a statistical technique to discover latent patterns in a dataset, is used as a data-reduction technique to see "if there is a set of factors that can explain the variation of the variables under study" (Hinton, 2004, p. 395). By observing how each variable "loads" each factor, it is possible to reveal particular factors which characterize different latent dimensions.

All survey answers were stored in a MySQL database and then exported to Microsoft Excel for data cleansing (i.e., removing typos, formatting the data in a standard way, controlling synonyms, etc.). The cleaned data then were imported into SAS 9 software for descriptive statistics and regression analysis.

Results and Discussion

The results are described based on an analysis of individual questions in the survey. Table 4 lists each question in both the abbreviated form and the original full text of the complete question. Each subsection details the results for the analysis of one survey question (except Questions 17 and 18, which were combined). In each case, the analysis was performed on the complete dataset ($N = 2,063$).

TABLE 4. Individual survey questions.

Question	Short name	Complete question
9	Reading hours	In a typical week, how many hours do you spend reading information relevant to your work?
11	Frequency of use for different resource types	How often do you use the following types of resources as a primary source of information for your research?
15	Articles retrieved from specific sources	In a typical month, how many articles did you retrieve to read from the following sources?
16	Searchers' confidence in search process	How confident are you that you are finding everything you should on your topic?
17 & 18	Whether scientists maintain a personal article collection or a bibliographic database	Do you maintain a personal article collection? Do you maintain a personal bibliographic database for print and/or electronic references?
17.1	Size of print-/electronic-article collection	Indicate the approximate size of both your print- and electronic-article collections
17.2	How often scientists use their article collection	How often do you utilize your personal article collection?
19	Percentage of papers from their article collection that have notes	Of articles you have in your personal article collection, on what percentage of them have you made some sort of notes?
20	Searching preference (print vs. electronic)	If given the option, how would you prefer to search for information?
21	Reading preference (print vs. electronic)	If given the option, how would you prefer to read retrieved information (journal articles, etc.)?
22	Frequencies of visiting libraries in person in the last 12 months	How often did you visit university libraries in person in the last 12 months?
24	Factors determining choices of journals for their publication	What determines your choice of journals for the publication of your work?
25	Preference between Google and the library's home page	Which interface would you rather use to begin your search process?

The complete dataset of the survey results is publicly available at <http://ils.unc.edu/bmh/isb/CompleteData>, and other researchers are encouraged to study it.

In addition to analyzing for correlations between questions, it also is useful to group by individual department to study similarities and differences in user behaviors across departments. This is important because academic libraries often group their services based on broader disciplinary groups (e.g., math and physics, or health sciences). In the past, this was important to facilitate the handling of physical materials; in today's STEM libraries, most content is licensed digital content, potentially allowing for grouping by different factors such as similarities in information searching, collecting, and use behaviors.

Reading Hours

Scientists were asked how many hours in a typical week they spent reading information relevant to their work. This attempts to capture a generalized sense of reading and includes reading content in multiple formats (books, web pages, journal articles, etc.). This is likely to be a larger number than reported in some of the more comprehensive surveys which asked only about time spent reading journal articles (Brown, 1999, Tenopir & King, 2002, Tenopir et al., 2004). According to this survey, the average number of hours spent reading per week was 11. This additionally may be affected by factors not analyzed in this study, such as the amount of research activity.

To examine the potential contributing factors to the total reading hours of academics, Poisson regression analysis was applied; the results are shown in Table 5. As shown in Table 5, *age* and *number_of_library_visits* had a significantly

positive, but rather minor, impact on reading hours. Compared with doctoral students, research staff/adjunct, and other personnel read less, approximately 73% of the amount read by doctoral students [This percentage is calculated from the exponential function of the estimated value (0.3158), which equals 0.73]. Researchers who had maintained a personal bibliographic database read significantly more hours per week (4.39 times) than did those who had not maintained one. While a correlation is indicated, the underlying causal factor may not be clear. For instance, it may be that heavy readers are more likely to need bibliographic software to manage their research paper collections. With the exception of the factors mentioned earlier, none of the remaining variables significantly impacted reading hours. Note that no significant interaction was found between department (discipline) and reading hours. This differs from the findings of Tenopir et al. (2004), in which medical scientists were found to read more as compared to scientists.

Individual department reading hours were calculated to study disciplinary differences. The distribution of reading hours was spread evenly over the range of 5.75 to 23.7 hr per week, and the groupings of departments by reading hours did not naturally align with departmental type (natural science, engineering, and medical science). Most departments reported reading around 10 to 15 hr per week. The heaviest readers included those in Electrical Engineering (23.7 hr per week) and Mathematics & Statistics (20 hr per week).

Frequency of Use for Different Resource Types

To identify the primary sources of information used by researchers and how frequently they used these resources, participants were asked to list how often they used the

TABLE 5. Poisson regression for parameters predicting reading hours.

Parameter	Estimate	%
Demographic		
age	0.010**	1.01
gender		
Male = 1; female = 0	0.057	1.06
Psychological		
confidence	0.022	1.02
Role-related		
position ^a		
Master's student	0.129	1.14
Doctoral student	0.000	1.00
Postgraduate/fellow	-0.082	0.92
Research staff/adjunct	-0.316**	0.73
Assistant professor	-0.023	0.98
Associate professor	0.054	1.06
Professor	-0.113	0.89
Others	-0.307**	0.74
Environmental context		
department_type ^b		
Natural science	0.000	1.00
Medical science	0.073	1.08
Engineering	0.083	1.09
distance_to_library	0.030	1.03
Information-searching behavior		
preferred_search_format		
Electronic = 1; print = 0	-0.027	0.97
number_of_library_visit	0.002**	1.00
search_page		
Google = 1; Library's home page = 0	0.019	1.02
Information-using behavior		
percentage_of_articles_with_notes	0.001	1.00
preferred_reading_format ^c		
Print	0.000	1.00
Electronic	0.004	1.00
Both ways	0.085	1.09
Information-collecting behavior		
maintain_article_collection		
Y = 1; N = 0	-0.056	0.95
maintain_bibliographic_database		
Y = 1; N = 0	1.480**	4.39
percentage_of_articles_in_bibliographic_database	-0.001	1.00

^aReference group is doctoral student. ^bReference group is natural science. ^cReference group is print.

*Significant at the 0.1 level. **Significant at the 0.05 level.

following: books, journals, preprints, attendance at conferences, conference proceedings, web pages, online databases, and personal communication. Journals, web pages, and personal communications were reported as used on a daily basis to support research activities. In addition, researchers read books monthly or weekly, attended conferences annually, and rarely used preprints. To reduce and classify the eight resources into major groups according to their frequency of use, exploratory factor analysis was performed. As shown in Table 6, three latent factors were extracted, and resources were divided into four clusters based on the loadings of each factor. Web pages, online databases, and personal communication were heavily loaded on the first factor, probably due to their high frequency of use. Attendance at conferences and conference proceedings are highly associated with Factor 2,

TABLE 6. Exploratory factor analysis for frequency of using types of resources.

	Factor 1	Factor 2	Factor 3
Books	0.301	0.193	0.252
Journals	0.120	0.130	0.489
Preprints	0.088	0.339	0.396
Attendance at conferences	0.215	0.516	0.141
Conference proceedings	0.154	0.541	0.190
Web pages	0.589	0.201	0.056
Online databases	0.485	0.043	0.228
Personal communications	0.513	0.183	0.063

which is probably related to the context of conferences (i.e., that they occur annually). Journals and preprints loaded on the third factor perhaps because they are similar in format, packaging, and distribution. It may be interesting to see if this changes as articles become individually published online; that is, whether their behavior then becomes more like that for web pages and members of the first factor group. Note that books were not heavily loaded on any of the three factors, and because they could not be explained by any single factor, they ended up as a separate cluster.

Books, journals, conference proceedings, and web pages were picked as the representatives of the four clusters for the follow-up analysis. Ordered logit regression models were built for the usage of each (Table 7). Generally, the academic position and department type had the strongest effect on scientists' frequencies of using these resources. For example, postgraduate students, as compared with doctoral students, used books much less frequently. Engineers read fewer books and journals, but reported using more conference materials and conference proceedings as compared with natural scientists. Medical scientists reported reading even fewer books than did engineering scientists. The results of this study agree with the findings of Brown (1999) and Tenopir et al. (2004) that journals are the primary source relied on by scientists from all surveyed disciplines. The study results disagree with the conclusion of Whitmire (2002) that there were fewer differences between the information-seeking behaviors in the life science versus the non-life-science disciplines. The disagreement might be attributable to Whitmire's study population of undergraduate students. We did find that medical scientists reported reading slightly more journals than did the natural scientists, as suggested by Tenopir et al., but the difference was not significant. In addition to academic position and department type, researchers' preference of search-interface type (i.e., Google page over library's home page) positively affected their frequencies of using web pages and conference proceedings. This finding is supported by the conclusion of Meho and Yang (2007) that Google stood out in its coverage of conference proceedings as compared to other searching tools. Demographics had little impact, except for the unexpected finding that males read books more often than do females.

TABLE 7. Ordered logit regression for frequency of using types of resources.

Parameter	Books Estimate	Journals Estimate	Conference proceedings Estimate	Web pages Estimate
Demographic				
age	0.024	-0.008	0.004	0.026
gender				
Male = 1; female = 0	0.647**	0.047	0.120	0.070
Psychological				
confidence	-0.100	0.033	0.050	-0.112
Role-related				
position ^a				
Assistant professor	0.366	0.207	0.115	0.090
Associate professor	-0.403	0.564	-0.075	0.038
Master's student	-0.081	-0.439	0.099	0.627
Others	-2.164**	-1.648*	-1.977**	-0.198
Postgraduate/fellow	-1.754**	0.471	-0.958	-0.358
Professor	-0.807	0.746	-0.068	-0.550
Research staff/adjunct	-0.978**	-1.430**	-0.142	0.511
Environmental context				
department_type ^b				
Engineering	-0.628**	-0.922**	0.766*	0.219
Medical science	-0.838**	0.359	0.017	0.212
distance_to_library	0.292*	-0.129	-0.018	0.107
Information-searching behavior				
preferred_search_format				
Electronic = 1; print = 0	0.069	-1.107	0.257	1.007
number_of_library_visit	0.016**	0.010*	0.002	0.002
search_page				
Google = 1;	0.166	-0.023	0.438*	1.271**
Library's home page = 0				
Information-using behavior				
reading_hours_per_week_per_week	0.016	0.095**	0.010	0.051
percentage_of_articles_with_notes	0.001	0.009*	-0.005	0.005
preferred_reading_format ^c				
Both ways	0.295	0.187	0.169	0.203
Electronic	0.654	-0.174	0.131	0.132
Information-collecting behavior				
maintain_article_collection				
Y = 1; N = 0	0.000	0.000	0.000	0.000
maintain_bibliographic_database				
Y = 1; N = 0	0.000	0.000	0.000	0.000
percentage_of_articles_in_bibliographic_database	-0.003	-0.003	0.009**	0.001

^aReference group is doctoral student. ^bReference group is natural science. ^cReference group is print.

*Significant at the 0.1 level. **Significant at the 0.05 level.

Articles Retrieved From Specific Sources

One survey question asked scientists to estimate the number of articles that they retrieved from 13 types of sources. Researchers indicated a strong preference for electronic sources over print ones since the top-four most retrieved sources were all in electronic formats. Electronic journals subscribed to by their library and electronic open-access journals were the two primary resources most utilized. Exploratory factor analysis was conducted to identify the degree of similarity in popularity for each source. As shown in Table 8, five factors were extracted. Factors 2, 3, and 4 neatly separate the materials from three different channels: colleagues, laboratory research group, and external sources, respectively. Factor 5 is possibly related to electronic journals

because it is loaded both by library-subscribed e-journals and open-access e-journals. This factor also explains the variation in popularity since both the library-subscribed e-journals and the open-access e-journals are the most popular sources, as indicated earlier. Factor 1 is more difficult to explain, as it loads both library-subscribed and personally subscribed journals in either print or electronic formats. The loading coefficients on each factor suggest that these latent factors are more strongly associated with information channels (e.g., laboratory, colleagues, library, and personal) rather than with formats (i.e., electronic, print). For example, people who retrieved more from an electronic laboratory-subscribed journal also are likely to retrieve more from a print lab-subscribed journal. Most recent research has indicated that convenience

TABLE 8. Exploratory factor analysis for types of information sources.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
[Electronic] Personally subscribed journal	0.854	−0.001	0.009	0.018	0.025
[Electronic] Lab-subscribed journal	0.045	0.020	0.642	0.018	0.070
[Electronic] Library-subscribed journal	0.212	−0.018	0.022	0.202	0.429
[Electronic] Open-access journal or institutional repository or digital library	0.028	0.010	0.167	0.125	0.337
[Electronic] (Author's) Web site	−0.016	0.021	−0.007	0.036	0.327
[Electronic] Personal digital library	0.040	0.075	0.006	0.329	0.235
[Electronic] Colleague's e-copy	0.038	0.932	0.084	0.079	0.096
[Print] Personally subscribed journal	0.872	0.012	0.077	−0.017	0.077
[Print] Lab-subscribed journal	0.068	0.105	0.647	−0.018	0.048
[Print] Library-subscribed journal	0.812	0.031	0.090	0.066	0.078
[Print] Copy of colleague's	−0.003	0.935	0.086	0.005	−0.023
Interlibrary loan	0.008	0.002	0.001	0.385	0.064
Document-delivery service	0.001	0.005	0.005	0.467	0.077

of the information is the most desirable feature (Liu, 2006; Tenopir, 2003). Our results seem to agree: An electronic format is clearly preferred (It is most convenient.); beyond that, the factor analysis suggests that primary differences occur across delivery channels.

A representative from each cluster was picked for the follow-up analysis. A Poisson regression model was built for each to identify the potential explanatory factors. The regression results are shown in Table 9. For demographic variables, gender played a more important role than did age in the number of articles retrieved from different sources. Specifically, males retrieved more papers from electronic journals (either personal or lab subscriptions), but retrieved fewer papers from colleagues and relied less on a document-delivery service. Academic rank also was found to have an effect. Senior professors retrieved more than did junior ones from both library-subscribed and personally subscribed e-journals, and junior professors retrieved more than did graduate students. Academic involvement seems to be the explanation for this, as suggested by Palmer (1991b)—that the desire for immersion in an information-rich environment and the preference for drawing information from a wide array of sources are commonly attributed to creative scientists. Information searchers willing to try new experiences (labeled “Activists” by Honey & Mumford, 1986) also are likely to publish more papers per year and believe serendipity plays an important role in problem solving (Palmer, 1991b). Distance to the library had a significant impact on the usage of a document-delivery service: Increasing the distance to the library increased the use of a document-delivery service. Information behaviors also played a role. Searchers who preferred the Google interface tended to share electronic copies with colleagues more than did searchers who preferred the library home-page interface. Researchers who preferred to read online only (This is a small group.) retrieved many more electronic documents, on average, from two sources—their library or colleagues—as compared to researchers who preferred to read print only or the combination of electronic and print. In addition, academic scientists who maintained

a personal article collection or a bibliographic database retrieved fewer articles from personal or lab sources.

Searching Preference (Print vs. Electronic)

Researchers were asked about their preferences for searching electronically versus traditionally (through print). The preference for searching online over searching through print media was overwhelming at 96.3%. It is not surprising that with the rapid development of the Internet, electronic journals, and digital libraries, more people preferred to search for information online due to its convenience and accessibility. Electronic materials have been predominant resources for academic researchers, especially in information searching (Hemminger et al., 2007; Liu, 2006). A small minority of people (3.7%) still prefer to search traditionally. Examining this small group of people shows that they are not much different from others in either their demographics or other information behaviors. Anticipated interactions with age or department were not found.

Studying departments showed that most of them were compromised people who preferred to search electronically. There were differences in departments, though. Some departments relied more heavily on particularly useful print resources. For example, in the ecology department, 50% reported preferring to search on print rather than electronically.

Reading Preference (Print vs. Electronic)

In addition to the preference for searching, scientists also were asked how they preferred reading information. To reduce possible confusion in the results, this study distinguishes between preference for reading and preference for search. While most participants preferred to search electronically, the majority indicated that they preferred a combination of electronic and print formats for reading, depending on the situation. Only a small percentage of academics read only electronically or only using print. This is inconsistent with

TABLE 9. Poisson regression for kinds of information sources.

Parameter	[Electronic] Personally subscribed journal Estimate	[Electronic] Laboratory- subscribed journal Estimate	[Electronic] Library- subscribed journal Estimate	[Electronic] Colleague's copy Estimate	Document- delivery service Estimate
Demographic					
age	0.005	0.004	-0.028	-0.026**	0.008
gender					
Male = 1; female = 0	0.345**	0.467*	0.128	-0.316*	-0.547*
Psychological					
confidence	0.003	-0.189	0.094	-0.117	0.022
Role-related					
position ^a					
Master's student	-0.036	-0.029	-0.145	0.244	0.588
Doctoral student	0.000	0.000	0.000	0.000	0.000
Postgraduate/fellow	0.663**	1.541**	0.532**	0.656**	-2.691
Research staff/adjunct	0.075	-0.790	0.105	-0.140	-0.363
Assistant professor	0.814**	-0.855	0.633**	-0.113	0.888**
Associate professor	0.792**	-0.933	0.703**	-0.076	1.122**
Professor	1.342**	0.316	1.246**	1.246**	0.021
Others	0.319	0.718	0.096	0.108	-1.796
Environmental context					
department_type ^b					
Natural science	0.000	0.000	0.000	0.000	0.000
Medical science	-0.160	-0.543**	-0.072	-0.203	-0.429
Engineering	-0.498	-1.294	-0.306	-0.154	0.058
distance_to_library	0.062	-0.230	-0.065	0.0249	0.415**
Information-searching behavior					
preferred_search_format					
Electronic = 1; print = 0	-0.041	-0.721	0.835	0.329	-0.739
number_of_library_visit	0.001	-0.003	0.000	-0.000	0.009**
search_page					
Google = 1; library's home page = 0	0.024	-0.171	0.173	0.329**	-0.017
Information-using behavior					
reading_hours_per_week_per_week	0.011	0.018	0.032**	0.034**	0.041**
percentage_of_articles_with_notes	0.005*	0.011**	0.000	-0.002	0.003
preferred_reading_format ^c					
Print	0.000	0.000	0.000	0.000	0.000
Electronic	-0.081	0.465	0.551**	0.576**	0.640
Both ways	-0.001	0.577	0.051	0.348	0.073
Information-collecting behavior					
maintain_article_collection					
Y = 1; N = 0	-0.520*	-1.849**	-0.188	0.510	23.238
maintain_bibliographic_database					
Y = 1; N = 0	1.835	-2.284**	0.456	22.015	21.486
percentage_of_articles_in_ bibliographic_database	-0.000	0.006	-0.001	0.004*	0.006

^aReference group is doctoral student. ^bReference group is natural science. ^cReference group is print.

*Significant at the 0.1 level. **Significant at the 0.05 level.

Liu (2006), who reported that nearly 90% of participants preferred paper as a reading medium to digital media. It also disagrees with Tenopir's (2003) conclusion that print remained the most popular medium for books and that e-book use was still in its very early stage. This appears to be changing with the more widespread reading of PDFs using free, high-quality viewers. In addition, the rapid growth of e-book readers may be influencing readers (Walsh, 2010).

When people search, browse, scan, or extract facts, they prefer electronic documents, but when they need to do in-depth reading, many still prefer to print out and then annotate printed documents. Annotating and highlighting are common reading activities in the printed environment, yet they are not smoothly integrated into online reading (O'Hara & Sellen, 1997). An ordered logit regression model was built with reading preference as the independent variable

and print-only as the reference group. Results indicate that having a larger number of annotations within a person's article collection has a small, but statistically significant, interaction with their reading choice. Having more annotations is associated with a lower likelihood of reading only electronically, as compared with print or the combination of print and electronic reading. No other factors, either contextual or behavioral, had a significant effect on the reading preferences of academics. Given the limitations of current electronic reading and annotation tools, most scientists still prefer both reading on paper and electronically.

There is great variance for the reading preference across departments. Most departments follow the norm of preferring the combination of reading electronically and on print. Print-only reading was more popular than electronic-only reading, but both of these were much less preferred than was a combination of the two methods. As exceptions, the Applied Sciences sample all preferred to read electronically while those in Aging Studies preferred to read through print media only.

Preference Between Google and the Library's Home Page

The vast majority of academic searching for research is now conducted using either library web sites or search engines. To identify which type of interface is preferred in a variety of academic fields, the survey asked participants to indicate their preference between the Google search interface and their library home-page search interface. Responses from participants were split nearly evenly between the two types of interfaces. To further examine the potential factors that affect the choice of interface, a logit regression model was constructed for the binary preference of the search interface; results are summarized in Table 10. The results indicate that males prefer searching the Google interface more [$\exp(0.5324) = 1.7$ times] than do females. This interesting gender difference might be explained by Burdick's (1996, p. 33) statement that men prefer something extensive and complete while women tend to "investigate and formulate." Compared with most library search pages, the Google search interface provides a comprehensive list of results from all kinds of sources, which are much more extensive and need further exploration to identify the most relevant information. On the other hand, women may prefer the better organization of results, prepared by authorities, of the library search-page interface. *Confidence* is slightly different between the two types of searchers. Google searchers are a bit less confident that they are finding everything they should. A possible explanation might be that "Google is too generic and cannot guarantee the relevance of the results it gives. . . ." (Vibert, Rouet, Ros, Ramond, & Deshouillieres, 2007). Another possible reason is that experienced searchers tend to be familiar with the library search interface and trust the authority of the content it provides, thus feeling more confident than does someone using a Google search where the searcher is inundated with a large set of search results without clear guidance as to the relevance of individual results. Research staff and

TABLE 10. Binary logit regression for parameters predicting the preference of search interface among scientists: Google versus the library home page.

Parameter	Estimate	%
Demographic		
age	-0.009	0.99
gender		
Male = 1; female = 0	0.532**	1.70
Psychological		
confidence	-0.135*	0.87
Role-related		
position ^a		
Master's student	-0.397	0.67
Doctoral student	0.000	1.00
Postgraduate/fellow	-0.031	0.97
Research staff/adjunct	1.101**	3.01
Assistant professor	-0.074	0.93
Associate professor	-0.406	0.67
Professor	-0.034	0.97
Others	0.118	1.12
Environmental context		
department_type ^b		
Natural science	0.000	1.00
Medical science	0.052	1.05
Engineering	0.076	1.08
distance_to_library	0.006	1.01
Information-searching behavior		
preferred_search_format		
Electronic = 1; print = 0	0.066	1.07
number_of_library_visit	-0.003	1.00
Information-using behavior		
reading_hours_per_week_per_week	0.004	1.00
Percentage_of_articles_with_notes	0.000	1.00
preferred_reading_format ^c		
Print	0.000	1.00
Electronic	0.336	1.40
Both ways	0.271*	1.31
Information-collecting behavior		
maintain_article_collection		
Y = 1; N = 0	-0.702**	0.50
maintain_bibliographic_database		
Y = 1; N = 0	0.909	2.48
percentage_of_articles_in_bibliographic_database	0.002	1.00

^aReference group is doctoral student. ^bReference group is natural science. ^cReference group is print.

*Significant at the 0.1 level. **denotes significant at the 0.05 level.

adjuncts are much more likely than are doctoral students to prefer the Google search interface. Scientists who maintain a personal article collection tend to use the library's home page rather than Google. This may be due to factors such as the academic nature of the integrated library system, which primarily contains academic articles rather than all types of formats retrieved by Google. Another factor might be the support provided by the library interface for directly importing articles into the user's personal article collection. The differences between departmental types for search interface preference were minor.

However, individual departmental differences with respect to search-page preference are notable. Approximately half of

the departments reported an evenly split preference (50%) between Google and the library home page. Geology reported the lowest preference for the Google interface (23.5%) whereas Applied Science, Education, the Lineberger Cancer Center, Natural History, Physical Therapy & Rehabilitation Sciences, and Urban Transportation Research reported the highest (100%).

Searchers' Confidence in the Search Process

Scientists were asked to rate their confidence level on a Likert scale of 1 (*very NOT confident*) and 5 (*very confident*) that they were finding everything that they should. Consistent with the central tendency bias when using the Likert scale, most of them chose a moderate score of 3 or 4 and avoided the extremes. To examine the influence of contextual factors on confidence level, ordered logit regression analysis was applied with the confidence level as the outcome variable (Table 11). Gender matters more than does age in affecting the confidence of researchers. On average, men had a higher confidence score than did women. This gender difference agrees with previous literature both from the information-seeking behavior research field as well as other fields. For example, Palmer (1991a) suggested that females may approach the information service with less confidence than do men. Possible contributing factors might be different computer or technology competency between the two genders (Jakobsdottir, Jonsdottir, & Hjartarson, 2004; Marley, 2007; Voelck, 2003) or just a reflection of the different psychological makeup of the two genders and not at all with real searching performance. Academic roles also make a difference, as professors and associate professors are more confident than are doctoral students. This might be explained by academic experience and searching competency. In addition, preference for search interface also makes a difference: Those who preferred the Google search interface tended to feel less confident about their search than did those who preferred the library's home page.

Confidence-level ratings were similar across departments, with most ratings being a 3 or 4 of 5. Two outliers are Physical Therapy & Rehabilitation Sciences, with very low ratings (2), and Education, with very high (4.5) ratings. The small sample sizes of these two departments may have contributed to this finding.

Scientists were asked whether they maintained a personal article collection; most responded that they did. The next question asked them to estimate the size of their print and electronic collections. Researchers' maintaining print collections of articles has been an established practice (Moore & McCabe, 1998) while keeping electronic journal article collections seems to be a newer practice. However, since 2000, the amount of material in electronic collections seems to be rapidly growing. Most people maintain collections in both formats based on their responses about the size of their collections. When asked whether they maintained a bibliographic database, over 50% researchers answered "no." Having a

TABLE 11. Ordered logit regression of scientists' confidence.

Parameter	Estimate	%
Demographic		
age	−0.022**	0.98
gender		
Male = 1; female = 0	0.374**	1.45
Role-related		
position ^a		
Master's student	−0.157	0.85
Doctoral student	0.000	1.00
Postgraduate/fellow	0.313	1.37
Research staff/adjunct	0.539**	1.71
Assistant professor	0.239	1.27
Associate professor	0.517**	1.68
Professor	0.828*	2.29
Others	0.410	1.51
Environmental context		
department_type ^b		
Natural science	0.000	1.00
Medical science	0.005	1.01
Engineering	−0.312	0.73
distance_to_library	0.102	1.11
Information-searching behavior		
preferred_search_format		
Electronic = 1; print = 0	−0.014	0.99
number_of_library_visit	0.001	1.00
search_page		
Google = 1; library's home page = 0	−0.215*	0.81
Information-using behavior		
reading_hours_per_week_per_week	0.007	1.01
percentage_of_articles_with_notes	0.003	1.00
preferred_reading_format ^c		
Print	0.000	1.00
Electronic	0.385*	1.47
Both ways	−0.030	0.97
Information-collecting behavior		
maintain_article_collection		
Y = 1; N = 0	−0.150	0.86
maintain_bibliographic_database		
Y = 1; N = 0	0.489	1.63
percentage_of_article_in_bibliographic_database	0.002	1.00

^aReference group is doctoral student. ^bReference group is natural science. ^cReference group is print.

*Significant at the 0.1 level. **Significant at the 0.05 level

personal bibliographic database means having not just a simple electronic-document collection but utilizing a piece of software such as EndNote and RefWorks to manage their collection with corresponding bibliographic information. To further examine the potential contributing factors to whether scientists have a personal article collection or a bibliographic database, logit regression was applied for the binary outcome. The regression results are summarized in Table 12. The results showed that participants from the medical sciences are less likely to maintain an article collection compared to their natural science participants. Searchers who preferred the Google interface were less likely to maintain an article collection compared to those who preferred their library's home-page interface. This might be due to the personality type of those preferring the library home page,

TABLE 12. Logit regression of whether scientists maintain a personal article collection.

Parameter	Estimate	%
Demographic		
age	0.020	1.02
gender		
Male = 1; female = 0	-0.371	0.69
Psychological		
confidence	-0.167	0.85
Role-related		
position ^a		
Master's student	0.963	2.62
Doctoral student	0.000	1.00
post graduate/fellow	-0.155	0.86
Research staff/adjunct	-0.224	0.80
Assistant professor	-0.596	0.55
Associate professor	-0.214	0.81
Professor	0.011	1.01
Others	0.462	1.59
Environmental context		
department_type ^b		
Natural science	0.000	1.00
Medical science	-0.654**	0.52
Engineering	1.078	2.94
distance_to_library	-0.056	0.95
Information-searching behavior		1.00
preferred_search_format		
Electronic = 1; print = 0	0.025	1.03
num_of_library_visit	0.001	1.00
search_page		
Google = 1; library's home page = 0	-0.721**	0.49
Information-using behavior		1.00
reading_hours_per_week_per_week	-0.013	0.99
percentage_of_articles_with_notes	0.015**	1.01
preferred_reading_format ^c		
Print	0.000	1.00
Electronic	0.888	2.43
Both ways	-0.596	0.55
Information-collecting behavior		1.00
maintain_bibliographic_database		
Y = 1; N = 0	-13.170	0.0
percentage_of_articles_in_bibliographic_database	0.009**	1.01

^aReference group is doctoral student. ^bReference group is science.

^cReference group is print.

*Significant at the 0.1 level. **Significant at the 0.05 level.

who may more carefully collect and maintain their materials. It also may be influenced by the varied formats of types of materials returned in Google searches, not all of which can easily be stored in academic article collections. On the other hand, for maintaining bibliographic databases, no significant contributing factors were found. It seems that use of bibliographic software is a personal behavior and is not strongly correlated with any contextual factors.

In all departments, at least 50% of the people maintained an article collection. The distribution falls roughly into thirds, with percentage ranges of 50 to 80, 80 to 95, and 100%, respectively. For maintaining a bibliographic database, there is much more variation across departments (range = 0–75%).

Size of Print/Electronic-Article Collection

In addition to the question of whether scientists have an article collection, we further asked them to list the number of items in their print and electronic-article collections. Generally speaking, for both print and electronic-article collections, the size follows a normal distribution, with a mean in range of 100 to 500 articles. Contributing factors are identified through an ordered logit model. According to Table 13, demographic variables, including both age and gender, contribute to the size of collection. Age was weakly positively associated with the print-collection size, and males indicated a larger electronic collection than did females. The gender difference possibly could be explained by the difference in digital technology acceptance for the two genders (Jakobsdottir et al., 2004; Marcella & Binfield, 1996; Marley, 2007). Academic roles also have a strong effect: Academic rank approximately demonstrates a co-increasing association with the size of print-article collection (i.e., senior professors have larger print-article collections than do junior professors; in turn, junior professors have larger print-article collections than do graduate students). A longer set of years for accumulating articles may be a good explanation for this result. Comparing by department type, medical science and engineering participants had smaller collections than did those in natural science. This is probably due to the different focuses for the different disciplines. Papers are more likely to be collected by scientists than they are by engineers and doctors. Note that researchers' preferred choice of search interfaces has the strongest impact on the collection size. Searching online corresponds with larger print and electronic collections, which could be attributable to the speed and convenience of finding vast amounts of information on the web. On the other hand, a preference for reading electronically correlates with smaller collections of print papers. Most people will save documents online if they read them online. Another point of interest is the impact of distance to the library. The farther the library, the larger the print collection the person has.

How Often Scientists Use Their Article Collection

The next question asked participants how frequently they used their article collection. More than half of the respondents said that they use it daily or weekly. An ordered logit regression model was built for the frequency level; results are shown in Table 14. Gender had a significant impact in that males used their article collection more often than did females. As to disciplines, those in engineering and medical science did not utilize their article collection as frequently as did those in natural science. This may be related to engineering and medical researchers investing less in their article collections (i.e., having a relatively small size of article collections). Generally, basic scientists had the most active information-collecting behavior, according to both the size of their collection and their frequency of using it. In addition, maintaining a bibliographic database was highly correlated

TABLE 13. Ordered logit regression of article-collection size (both print and electronic).

Parameter	Print		Electronic	
	Estimate	%	Estimate	%
Demographic				
age	0.054**	1.06	0.008	1.01
gender				
Male = 1; female = 0	0.031	1.03	0.513**	1.67
Psychological				
confidence	0.073	1.08	0.032	1.03
Role-related				
position ^a				
Assistant professor	0.930**	2.54	0.712**	2.04
Associate professor	1.029**	2.80	0.368	1.45
Master's student	-0.863**	0.42	-0.233	0.79
Others	0.294	1.34	0.313	1.37
Postgraduate/fellow	0.411*	1.51	0.691**	2.00
Professor	1.067**	2.91	0.915**	2.50
Research staff/adjunct	0.535*	1.71	0.178	1.20
Environmental context				
department_type ^b				
Engineering	-1.224**	0.29	-0.326	0.72
Medical science	-0.463**	0.63	-0.182	0.83
distance_to_library	0.170**	1.19	0.091	1.10
Information-searching behavior				
preferred_searching_format				
Electronic = 1; print = 0	0.764*	2.15	1.015**	2.76
number_of_library_visit	-0.000	1.00	-0.001	1.00
search_page				
Google = 1; library's home page = 0	-0.380**	0.68	-0.178	0.84
Information-using behavior				
reading_hours_per_week	0.021**	1.02	0.036**	1.04
percentage_of_articles_with_notes	0.003	1.00	-0.003	1.00
preferred_reading_format ^c				
Both ways	-0.592**	0.55	0.063**	1.06
Electronic	-0.618**	0.54	0.544	1.72
Information-collecting behavior				
maintain_collection				
Y = 1; N = 0	15.546	0.56	0.764	2.15
maintain_bibliographic_database				
Y = 1; N = 0	16.195	0.18	13.503	0.73
percentage_of_articles_in_bibliographic_database	0.003	1.00	0.004**	1.00

^aReference group is doctoral student. ^bReference group is science. ^cReference group is print.

*Significant at the 0.1 level. **Significant at the 0.05 level.

with the frequency of using the article collection, probably because these are closely related information-collecting behaviors. Note that reading hours and annotating behavior had a very small, but significant, positive impact on the usage frequency.

Percentage of Papers From Their Article Collection That Have Notes

This question investigated annotation behaviors by asking researchers how many articles in their collection they had annotated. Annotation is related to information retrieval and information searching. For example, annotation could be suggested as implicit indicators of document relevance (Kelly,

2004). Currently, more convenient tools for annotating or collaboratively annotating electronic documents are needed (Hemminger, Lu, Vaughan, & Adams, 2007). In this study, approximately 35% of the papers in the article collections had annotations. To identify the underlying correlation with contextual factors, a Poisson regression model was built for the percentage of annotations; results are summarized in Table 15. The model shows that demographic variables, including both age and gender, had significant contributions, although the impact of age was minor. Males do not like to annotate their papers as much as do females. Different annotating habits between the two genders might be due to different reading habits such as skimming or perusing. Academic position also has an impact, as professors do not tend to annotate

TABLE 14. Ordered logit regression of frequency for using article collection.

Parameter	Estimate	%
Demographic		
age	0.006	1.01
gender		
Male = 1; female = 0	0.294**	1.34
Psychological		
confidence	0.098	1.10
Role-related		
position ^a		
Master's student	-0.852**	0.43
Doctoral student	0.000	1.00
Postgraduate/fellow	0.128	1.14
Research staff/adjunct	-0.459	0.63
Assistant professor	0.134	1.14
Associate professor	0.131	1.14
Professor	-0.290	0.75
Others	-0.946**	0.39
Environmental context		
department_type ^b		
Natural science	0.000	1.00
Medical science	-0.233*	0.79
Engineering	-0.441*	0.64
distance_to_library	-0.015	0.98
Information-searching behavior		
preferred_searching_format		
Electronic = 1; print = 0	0.529	1.70
number_of_library_visit	-0.000	1.00
search_page		
Google = 1; library's home page = 0	-0.027	0.97
Information-using behavior		
reading_hours_per_week	0.043**	1.04
percentage_of_articles_with_notes	0.009**	1.01
preferred_reading_format ^c		
Print	0.000	1.00
Electronic	0.167	1.18
Both ways	0.212	1.24
Information-collecting behavior		
maintain_collection		
Y = 1; N = 0	0.000	1.00
maintain_bibliographic_database		
Y = 1; N = 0	2.264**	9.62
percentage_of_articles_in_bibliographic_database	0.003*	1.00

^aReference group is doctoral student. ^bReference group is science.

^cReference group is print.

*Significant at the 0.1 level. **Significant at the 0.05 level.

their articles as much as do students. This may be because students mark up more content because they have less experience or because of other factors such as different reading purposes. In addition to demographic and role-related variables, the preferred way of both searching and reading articles greatly affects people's annotation behavior. Both searching online and reading online significantly reduced the likelihood of making annotations, which is probably due to the lack of easy-to-use annotation tools supporting skimming or reading large amounts of electronic articles. Also note that those with personal article collections or with more articles in their bibliographic databases tended to have more papers with annotations.

TABLE 15. Poisson regression of percentage of papers with notes from their article collection.

Parameter	Estimate	%
Demographic		
age	-0.009**	0.99
gender		
Male = 1; female = 0	-0.314**	0.73
Psychological		
confidence	0.046	1.05
Role-related		
position ^a		
Master's student	-0.009	0.99
Doctoral student	0.000	1.00
Postgraduate/fellow	-0.067	0.94
Research staff/adjunct	-0.175	0.84
Assistant professor	-0.190**	0.83
Associate professor	-0.376**	0.69
Professor	-0.377**	0.69
Others	-0.047	0.95
Environmental context		
department_type ^b		
Natural science	0.000	1.00
Medical science	0.077	1.08
Engineering	0.107	1.11
distance_to_library	-0.073**	0.93
Information-searching behavior		
preferred_searching_format		
Electronic = 1; print = 0	-0.251*	0.78
number_of_library_visit	-0.000	1.00
search_page		
Google = 1; library's home page = 0	0.004	1.00
Information-using behavior		
reading_hours_per_week	0.004	1.00
preferred_reading_format ^c		
Print	0.000	1.00
Electronic	-0.185**	0.83
Both ways	-0.034	0.97
Information-collecting behavior		
maintain_collection		
Y = 1; N = 0	0.386**	1.47
maintain_bibliographic_database		
Y = 1; N = 0	0.849	2.34
percentage_of_articles_in_bibliographic_database	0.004**	1.00

^aReference group is doctoral student. ^bReference group is science.

^cReference group is print.

*Significant at the 0.1 level. **Significant at the 0.05 level.

Except for a few outliers, the percentage of annotations on papers by department was fairly uniform, and ranged from 20 to 60%, with most averaging around 40%.

Frequency of Visiting Libraries in Person in the Last 12 Months

There has been an already-documented decline in the number of visits to the library by academic researchers (Applegate, 2008; Gayton, 2008; Hiller, 2002; Odlyzko, 2002). In our study, researchers were asked how many times they has visited the library in person during the last 12 months; the average number of library visits was 25. The number of visits follows a power law distribution, with a

TABLE 16. Poisson regression of frequency for visiting libraries.

Parameter	Estimate	%
Demographic		
age	0.018**	1.02
gender		
Male = 1; female = 0	0.441**	1.55
Psychological		
confidence	0.032	1.03
Role-related		
position ^a		
Master's student	0.632**	1.88
Doctoral student	0.000	1.00
Postgraduate/fellow	-0.207	0.81
Research staff/adjunct	-0.691**	0.50
Assistant professor	-0.786**	0.46
Associate professor	-1.277**	0.28
Professor	-1.372**	0.25
Others	-0.302	0.74
Environmental context		
department_type ^b		
Natural science	0.000	1.00
Medical science	-0.315**	0.73
Engineering	-0.161	0.85
distance_to_library	-0.310**	0.73
Information-searching behavior		
preferred_searching_format		
Electronic = 1; print = 0	-0.184	0.83
search_page		
Google = 1; library's home page = 0	-0.218**	0.80
Information-using behavior		
reading_hours_per_week	0.026**	1.03
percentage_of_articles_with_notes	-0.000	1.00
preferred_reading_format ^c		
Print	0.000	1.00
Electronic	0.039	1.04
Both ways	0.094	1.10
Information-collecting behavior		
maintain_collection		
Y = 1; N = 0	0.074	1.08
maintain_bibliographic_database		
Y = 1; N = 0	1.655	5.23
percentage_of_articles_in_bibliographic_database	-0.001	1.00

^aReference group is doctoral student. ^bReference group is science.

^cReference group is print.

*Significant at the 0.1 level. **Significant at the 0.05 level.

minority of frequent visits and a long tail of very few visits. To identify potentially correlated factors, a Poisson regression was applied to the visit frequencies, as shown in Table 16. Our results indicate that demographic variables play an important contributing role. Gender matters more than does age, and males unexpectedly visit the library more often than do females. This may be partly explained by Karim and Hasan's (2007) conclusion that men read more resources than do females other than books, including newspapers, and literature. With respect to age, these results disagree with Rowland and Nicholas's (2008) report that as age increases, visits to the library gradually decreases. In this study, age is a rather weak predictor since academic positions are considered in the same model which "absorbs" some of the effect of

age. In addition to demographic variables, academic position has a strong effect on the number of library visits. Academic rank seems negatively associated with the frequency of visits. The number of library visits for professors is only one third as many as those for master's students. It is interesting that those who prefer the Google search interface do not visit the library as often as do searchers who prefer the library home-page interface.

For individual departments, 20 visits per year is the median, with the distribution ranging from 0 to 60 visits. The one exception was Information & Library Science, with 122 visits on average. This was likely due to the sample consisting only of participants from UNC, where the library physically resides in the department itself.

Factors Determining Choices for Journals of Their Publication

The survey listed nine possible criteria that scientists might use when choosing a particular journal for publishing their work. Participants were asked to rate the criteria on a Likert scale of 1 (*very insignificant*) to 5 (*very significant*). The audience and standing of journal were the highest rated criteria. Exploratory factor analysis was conducted to examine the common variance shared by these nine criteria. As shown in Table 17, three factors are extracted. Ability to include links, color, and so on, and availability on campus were heavily loaded on the first factor, which might be related to "features" of the journal separate from the content and intended audience. The second factor could be related to author publication costs because it includes two factors, the author having to pay to publish and author page charges. Audience, standing, and speed were loaded in the third factor, which all relate to impact.

"Support to open access to journal articles," "author having to pay cost of publication," and "standing of journal in your field" were picked as the representatives of the three clusters for the follow-up regression analysis. As shown in Table 18, academic position had the strongest effects. For example, professors tended to give a lower score for publishing cost, but a higher score on the journal standing as compared to doctoral students' ratings. This may be due to professors being less concerned about cost and more concerned about quality of the journal. Demographic variables of gender and age had the next-largest effects. Gender played a surprising role: Males cared less about open access than did females. Age had a small effect related to publishing cost and journal standing. Participants who preferred to read electronically only or in electronic and print formats cared more about open access. This may be explained by their higher level of acceptance of content in an electronic format. Author paying publication costs was related to several information behaviors including number of library visits, search page interface, reading hours, and maintaining a bibliographic database. Note that searchers who preferred the Google interface and bibliographic database maintainers both cared less about having to pay the cost of publication. Department type does not have

TABLE 17. Exploratory factor analysis for choosing journals for publication.

	Factor 1	Factor 2	Factor 3
Ability to include links, color, graphics, multimedia	0.454	0.212	0.062
Audience	-0.079	0.068	0.673
Author having to pay cost of publication	0.223	0.725	0.135
Availability on campus	0.580	0.191	-0.022
Editorial board	0.283	0.167	0.350
Page charges for long articles	0.363	0.727	0.069
Speed of publication	0.303	0.252	0.422
Standing of journal in your field	0.047	-0.011	0.656
Support of open access to journal articles	0.594	0.133	0.088

TABLE 18. Ordered logit regression of reasons for choosing journals.

Parameter	Support open access for journal articles Estimate	Author having to pay cost of publication Estimate	Standing of journal in your field Estimate
Demographic			
age	0.008	0.020**	-0.040**
gender			
Male = 1; female = 0	-0.246**	-0.056	0.141
Psychological			
confidence	0.108*	0.037	0.052
Role-related			
position ^a			
Master's student	0.473**	0.094	-1.220**
Doctoral student	0.000	0.000	0.000
Postgraduate/fellow	0.193	-0.056	0.176
Research staff/adjunct	0.019	0.442	-0.352
Assistant professor	-0.279	-0.458**	1.172**
Associate professor	-0.516**	0.120	0.898**
Professor	-0.384	-0.491*	1.627**
Others	-0.217	-0.986**	-0.038
Environmental context			
department_type ^b			
Natural science	0.000	0.000	0.000
Medical science	-0.088	-0.175	0.119
Engineering	0.134	-0.124	-0.601
distance_to_library	0.062	0.082	0.061
Information-searching behavior			
preferred_searching_format			
Electronic = 1; print = 0	-0.325	0.507	-0.145
number_of_library_visit	0.001	0.004**	-0.002
search_page			
Google = 1; library's home page = 0	-0.004	-0.297**	0.309
Information-using behavior			
reading_hours_per_week	0.007	0.020**	-0.000
percentage_of_articles_with_notes	0.002	0.000	0.003
preferred_reading_format ^c			
Print	0.000	0.000	0.000
Electronic	0.356*	0.186	-0.095
Both ways	0.427**	0.069	-0.110
Information-collecting behavior			
maintain_collection			
Y = 1; N = 0	-0.119	0.354	0.433
maintain_bibliographic_database			
Y = 1; N = 0	1.223	-1.697*	-0.132
percentage_of_articles_in_bibliographic_database	-0.0013	-0.000	0.000

^aReference group is doctoral student. ^bReference group is science. ^cReference group is print.

*Significant at the 0.1 level. **Significant at the 0.05 level.

an effect on preference of journals for publishing. This is important because it suggests that the criteria used for choosing journals in which to publish are commonly valued among disciplines and that there is a core set of reasons applicable to all fields.

Conclusion and Limitations

In this article, a framework for understanding the information-seeking behavior of scientists was built, and regression techniques were applied to the survey data of 2,063 academic researchers in natural science, engineering, and medical science from five research universities in the United States. This study explored the influence of demographics, environmental context, and disciplines on the information behavior of scientists. Overall, many factors were found to affect the specific information-seeking behaviors of scientists, including demographic, psychological, role-related, and environmental factors. Of the factors having an effect, academic position was the most important determinant of information behavior; this might be due to the academic setting of this study. Other factors that often played a role were gender and discipline. While there were differences at the discipline (i.e., department) level and at the department-type level (i.e., medical, natural science, engineering), these were not consistent across information behaviors. Thus, it is difficult to recommend different support strategies based on department groupings or department types. Information-seeking behavior could be said to be conditioned by what library and information services were available, such as distance to the library, available electronic databases, and so on. Factor analysis was used to successfully reduce the quantity of data and to provide a concise and understandable account of information behavior. Beyond the factors mentioned earlier, information behavior demonstrates individual differences that could not be easily explained by the factors examined in this study. These findings agree with the work of Palmer (1991b), who found that information behavior seemed independent of discipline and more determined individually as a naturally enjoyed activity, and Davies (1998), who concluded that the information-seeking and handling habits of researchers are very personal and thus may be less affected by external factors.

There are some limitations to this study. First, the framework is a simplified and high-level abstraction of the relationships to accommodate the existing survey variables. As such, it might fail to capture some hidden factors that may potentially contribute to explaining the behavior of academic scientists, such as training received in graduate schools which may affect later behavior. Second, the regression model reveals correlated relationships between variables, but not causal relationships. Thus, this is an exploratory study rather than an explanatory one, and the regression analyses cannot answer questions such as "Why do medical scientists read fewer books than do natural scientists?" Third, technical issues related to regression analysis, such as missing data issues, goodness of fit, and correlations of independent

variables were not considered in the regression model for simplification purposes. While this is common practice and not usually a problem for a dataset of this size, different handling strategies might lead to different results. Fourth, data collected from a survey might provide an oversimplified and superficial understanding of this complex human behavior. To better understand the underlying motivation of information-seeking behavior among scientists, a complementary study using in-depth interviews combined with information behavior captured through screen logging and diaries is being conducted. In addition, work is being done to perform longitudinal comparisons to study the trend of scientists' information-seeking behavior over the past several decades, with the aim of more clearly understanding long-term changes in the information-seeking behavior of scientists from a historical perspective.

Acknowledgment

This work was supported by a grant from the Ochiltree Foundation Grant ("Information Seeking Behavior of Scientists," Bradley Hemminger, Principal Investigator).

References

- Applegate, R. (2008). Whose decline? Which academic libraries are "deserted" in terms of reference transactions? *Reference and User Services Quarterly*, 48(2), 176–189.
- Boneva, B., & Kraut, R. (2002). Email, gender, and personal relationships. *The Internet in Everyday Life* (pp. 372–403). Malden, MA: Wiley-Blackwell.
- Brown, C.M. (1999). Information seeking behavior of scientists in the electronic information age: Astronomers, chemists, mathematicians, and physicists. *Journal of the American Society for Information Science and Technology*, 50(10), 929–943.
- Brynin, M. (2006). *The neutered computer. Computers, Phones, and the Internet: Domesticating Information Technology*. New York: Oxford Press.
- Buente, W., & Robbin, A. (2008). Trends in internet information behavior, 2000–2004. *Journal of the American Society for Information Science and Technology*, 59(11), 1743–1760.
- Burdick, T.A. (1996). Success and diversity in information seeking: Gender and the information search styles model. *School Library Media Quarterly*, 25(1), 19–26.
- Davies, M. (1998). Impact of information developments on the information handling techniques of research scientists: Implications of selected results. *New Review of Information Networking*, 11, 53–70.
- Dillon, I.F., & Hahn, K.L. (2002). Are researchers ready for the electronic-only journal collection? Results of a survey at the University of Maryland. *Portal: Libraries and the Academy*, 2(3), 375–390.
- Gayton, J.T. (2008). Academic libraries: "Social" or "communal?" *The Nature and Future of Academic Libraries. Journal of Academic Librarianship*, 34(1), 60–66.
- Hargittai, E. (2002). Second-level digital divide: Differences in people's online skills. *First Monday*, 7(4), 1–20.
- Hemminger, B.M., Lu, D., Vaughan, K., & Adams, S.J. (2007). Information seeking behavior of academic scientists. *Journal of the American Society for Information Science and Technology*, 58(14), 2205–2225.
- Hiller, S. (2002). How different are they? A comparison by academic area of library use, priorities, and information needs at the University of Washington. *Issues in Science and Technology Librarianship*, 33.
- Hinton, P.R. (2004). *Statistics explained*. London, England: Routledge.
- Honey, P., & Mumford, A. (1986). *The manual of learning styles* (2nd ed.). Maidenhead, UK: P. Honey.

- Howard, P.E.N., Rainie, L., & Jones, S. (2002). Days and nights on the internet. *The Internet in Everyday Life* (pp. 45–73). Malden, MA: Wiley-Blackwell.
- Jakobsdottir, S., Jonsdottir, B., & Hjartarson, T. (2004). Gender, ICT-related student skills, and the role of a school library in an Icelandic school. *School Libraries WorldWide*, 10, 52–72.
- Karim, N.S.A., & Hasan, A. (2007). Reading habits and attitude in the digital age. *The Electronic Library*, 25(3), 285–298.
- Kelly, D. (2004). Understanding implicit feedback and document preference: A naturalistic user study. Unpublished doctoral dissertation, Rutgers University, New Brunswick, NJ.
- Kennedy, T., Wellman, B., & Klement, K. (2003). Gendering the digital divide. *IT & Society*, 1(5), 72–96.
- King, D.W., Tenopir, C., Montgomery, C.H., & Aerni, S.E. (2003). Patterns of journal use by faculty at three diverse universities. *D-Lib Magazine*, 9(10), 1–10.
- Liu, Z. (2006). Print vs. electronic resources: A study of user perceptions, preferences, and use. *Information Processing & Management*, 42(2), 583–592.
- Losh, S.C. (2004). Gender, educational, and occupational digital gaps 1983–2002. *Social Science Computer Review*, 22(2), 152.
- Marcella, R., & Binfield, S.J. (1996). Gender differences in the oral communication of technical information. *Education for Information*, 14(3), 181–193.
- Marley, J.L. (2007). Gender differences and distance education: Major research findings and implications for LIS education. *Journal of Education for Library and Information Science*, 48(1), 13–20.
- Meho, L.I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105.
- Moore, D.S., & McCabe, G.P. (1998). *Introduction to the practice of statistics*. New York, NY: Freeman.
- Mullis, R.L., Mullis, A.K., & Cornille, T.A. (2007). Relationships between identity formation and computer use among black and white emerging adult females. *Computers in Human Behavior*, 23(1), 415–423.
- Nelder, J., & Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A (General)*, 135(3), 370–384.
- Niu, X., Hemminger, B.M., & Lown, C. (2010). National Study of Information Seeking Behavior of Academic Researchers in the United States. *Journal of the American Society for Information Science and Technology*, 61(5), 869–890.
- Odlyzko, A. (2002). The rapid evolution of scholarly communication. *Learned Publishing*, 15, 7–19.
- O'Hara, K., & Sellen, A. (1997). A comparison of reading paper and on-line documents. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI '97)* (pp. 335–42). New York: ACM Press.
- Palmer, J. (1991a). Scientists and information: I. Using cluster analysis to identify information style. *Journal of Documentation*, 47(2), 105–129.
- Palmer, J. (1991b). Scientists and information: II. Personal factors in information behavior. *Journal of Documentation*, 47(3), 254–275.
- Pew Internet & American Life Project. (2000). Tracking online life: How women use the internet to cultivate relationships with family and friends.
- Rainie, L. (2000). Tracking online life: How women use the internet to cultivate relationships with family and friends. Pew Internet & American Life Project.
- Raban, Y., & Brynin, M. (2006). Older people and new technologies. *Computers, Phones, and the Internet: Domesticating Information Technology*. New York: Oxford University Press.
- Robbin, A., & Buente, W. (2008). Trends in internet information behavior, 2000–2004. *Journal of the American Society for Information Science and Technology*, 59(11), 1743–1760.
- Rowland, I., & Nicholas, D. (2008). Information behaviour of the researcher of the future. *Scholarly Communication Report*, 12(1), 7.
- Tenopir, C. (2003). Use and users of electronic library resources: An overview and analysis of recent research studies. Report of the Council on Library and Information Resources. Washington, DC: Council on Library and Information Resources.
- Tenopir, C., & King, D.W. (2002). Reading behaviour and electronic journals. *Learned Publishing*, 15(4), 259–265.
- Tenopir, C., King, D.W., & Bush, A. (2004). Medical faculty's use of print and electronic journals: Changes over time and in comparison with scientists. *Journal of the Medical Library Association*, 92, 233–241.
- Vibert, N., Rouet, J.F., Ros, C., Ramond, M., & Deshouillieres, B. (2007). The use of online electronic information resources in scientific research: The case of neuroscience. *Library and Information Science Research*, 29(4), 508–532.
- Voelck, J. (2003). Directive and connective: Gender-based differences in the management styles of academic library managers. *Portal: Libraries and the Academy*, 3(3), 393–418.
- Walsh, J. (2010). E-books: The end of the word as we know it. Retrieved from <http://www.independent.co.uk/arts-entertainment/books/features/is-the-publishing-industry-doomed-2099796.html>
- Whitmire, E. (2002). Disciplinary differences and undergraduates' information-seeking behavior. *Journal of the American Society for Information Science and Technology*, 53(8), 631–638.
- Wilson, T. (1997). Information behaviour: An interdisciplinary perspective. *Information Processing & Management*, 33(4), 551–572.