

Navigating Online Menus: A Quantitative Experiment

Karen Eliassen, Jill McKinstry, Beth Mabel Fraser, and Elizabeth P. Babbitt

A uniform interface to multiple, varied databases enhances searching and reduces the need for end-user training. The uniformity, however, can make it more difficult to select appropriate resources. This article describes a quantitative study at the University of Washington Libraries that examined the effect of terminology and screen layout on students' ability to correctly select databases from an introductory screen, the Session Manager. Results indicate a significant improvement in students' ability to navigate menu screens where terminology is expanded and selections are grouped by type. Implications for other projects, including Web design decisions, are suggested.

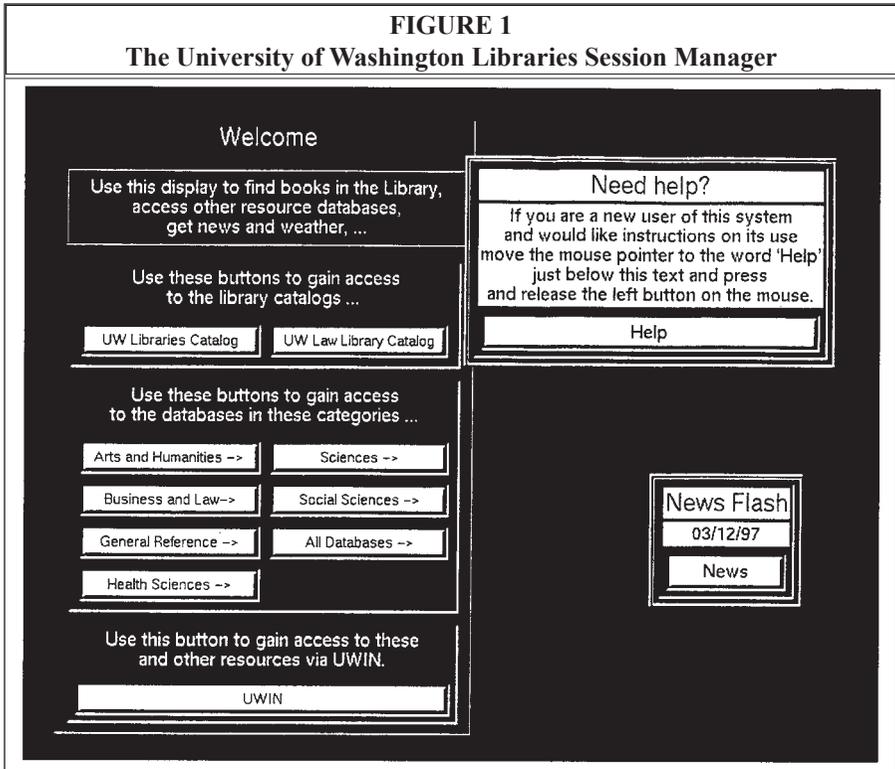


As online information proliferates, library users have more difficulty choosing among electronic resources. Welcome screens or menus help users make these decisions, but like all new tools, they require constant, iterative refinement in order to remain useful. This article reports the results of an empirical study at the University of Washington (UW) Libraries that examined two variables affecting the efficient navigation of the libraries' initial online interface, the Session Manager. Results are evaluated and used as a framework for further investigation. Implications for other projects, including Web-based design decisions, also are suggested.

Background

Willow, the uniform graphical user interface, is the hallmark of online searching at UW.¹ The single-search template and the ease with which a single query can be searched in multiple informational databases is popular with library users and has reduced the need for individual database training. Although the uniformity in searching and database appearance has made it easier to access multiple databases and to interpret results, it has made it more difficult to distinguish among electronic resources and to guide users in database selection. Although the Willow interface was a welcome alternative to a separate catalog and database interface, it

Karen Eliassen is Access Services/Reference Librarian in the Undergraduate Library at the University of Washington Libraries; e-mail: eliasen@u.washington.edu. Jill McKinstry is in the Networked Information Library at the University of Washington Libraries; e-mail: jillmck@u.washington.edu. Beth Mabel Fraser is Universal Access Project Librarian in the Reference and Research Services Division of the University of Washington Libraries; e-mail: bamf@u.washington.edu. Elizabeth P. Babbitt is Reference Librarian at Bellevue Regional Library in the King County (Washington) Library System; e-mail: ebabbitt@kcls.org.



was soon evident that for walk-up public terminal use, an introductory screen or program launcher was needed to help navigate in a graphical environment and to describe databases available from that terminal. The Session Manager was created to serve this function (see figure 1).

The arrangement of databases on the Session Manager was not a problem initially because there were only three databases: UW Libraries Catalog, MEDLINE, and ERIC. Each database had a separate button. But as databases were added, limited screen space required creating a hierarchy and grouping databases under a single arrangement. The design that evolved maintained separate buttons for the catalogs but grouped the bibliographic databases by subject. By 1997, the number of databases provided through the uniform Willow interface had grown to more than thirty.

Problem Statement

With the addition of so many new databases to the campus online system, many students were having difficulty locating the database they needed. At the same time, the role of the Session Manager had evolved. The increased importance of the Session Manager as a selection tool made it a part of the navigation process itself. Most users knew about MEDLINE and ERIC, but fewer knew what they would find in Avery, INSPEC, or Grolier's. Librarians reported users having problems ranging from misinterpretation of the information presented to a lack of understanding of library concepts, such as the difference between a catalog database and an index database. Typical complaints reported by the staff included these three examples:

1. "I have answered many, many questions at the information desk from people

TABLE 1
Research Design Showing Variables and Their Levels

Variable Type	Variable Name(s)	Levels or Attributes
Independent	(1) Terminology (2) Layout/grouping	(1) Existing terms, new terms (2) Existing layout, new layout
Moderator or Co-variate	Training level	Library instruction (yes/no)
Dependent	Success in selecting databases from session manager	Scores on a 15-point identification test
Control	170 undergraduate students	

who are using Books in Print to look for UW Libraries’ holdings. They just don’t seem to be able to tell the difference between the catalog and Books in Print.”

2. “About half of the questions and misunderstandings we address are based on confusion of the UW catalog with the various computerized indexes. We often have to tell students that the citations they have printed out have no direct relationship to UW holdings, or that they will not find a book on the shelves by browsing the indexes.”

3. “Today, a student wanted to look up books in the UW libraries. Because she was looking for books in the ‘Arts and Humanities,’ that is what she chose in the main menu.”

The first example shows the difficulty students have in differentiating between the content of the two databases, Books in Print and the UW Libraries Catalog. Librarians speculated that students looking for a book latched onto the term *books*, not listed anywhere else in the Session Manager. The second example illustrates a similar problem. Students have difficulty understanding the relationship between an abstracting and indexing database and the library catalog. The third example indicates the problem that arises from a subject arrangement of databases. Students looking for UW Libraries books on social science topics might choose the Social Science Databases button instead

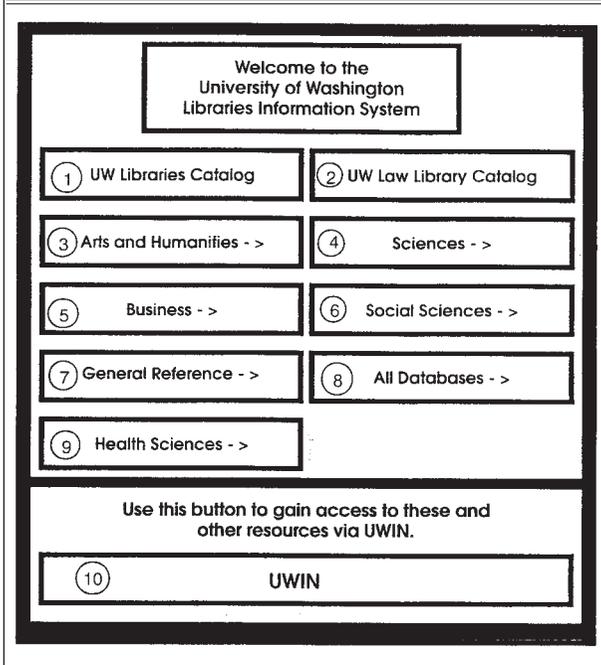
of the UW Libraries Catalog. In another typical example of confusion, students expect the All Databases button to allow them to search all the libraries’ databases simultaneously. That seems logical, but in fact this button merely takes them to a list of all databases. Librarians were convinced that the Session Manager could be altered in a way that would prevent students from making such mistakes.

The anecdotal information made it clear that the Session Manager often provided little navigation assistance to students. Along with the complaints came many suggestions from library staff for changes to the Session Manager. However, the stories did not give much insight into what changes would be most effective. Reluctant to use patrons as guinea pigs with a constantly changing interface, the libraries decided to gather information that would give a clearer direction for improvement. The Session Manager Evaluation Committee was created to collect data on how students interact with the Session Manager.

Design of the Session Manager Study

After reviewing the current literature related to the problem and meeting with other public services librarians to gather more details and hypotheses, the committee began the process of designing an experiment and creating a questionnaire to test hypotheses about the Session Man-

FIGURE 2
Low-fidelity Prototype of Existing Session Manager, Version A



ager.² Most of the perceived problems of navigation could be grouped into two large categories: terminology and layout. With an eye toward future interface evolution, the group decided to focus on these two issues of underlying significance. This decision guided development of the questionnaire and the research method. In designing the research outline, the committee concentrated on a single problem statement: What is the effect of terminology and screen layout/grouping on students' ability to correctly select databases from a Session Manager? Table 1 shows the variables and levels defined and identified by the group.

The committee created low-fidelity paper prototypes of the Session Manager. In order to judge the effects of terminology, the information on each button was increased to become more descriptive of the content. The text surrounding the

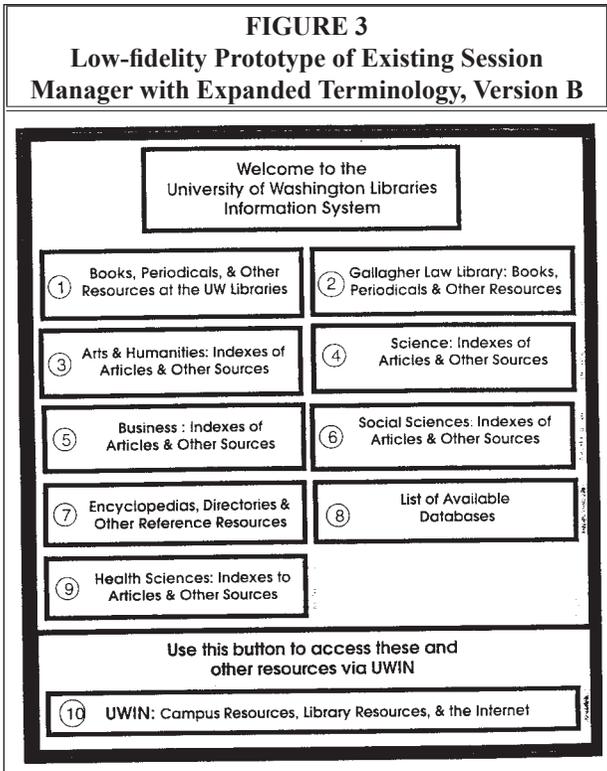
buttons also was altered to be more descriptive. To test layout/grouping, information on the initial selection screen was grouped according to type of information (catalogs, indexes, etc.) rather than subject category. In addition, training level was introduced as a variable in order to judge the impact of library instruction on participants. A clearly quantifiable dependent variable was chosen: the students' success in selecting databases from the libraries' Session Manager, where "success" was determined by scores on a fifteen-point identification test.

The committee also tried to anticipate threats to design validity in a way that would help to avoid common statistical errors. Similar empirical study designs were reviewed to judge, for example, whether the low-fidelity paper prototypes would provide an appropriate test instrument for an online interface. To ensure internal validity, the committee selected questions that would test the use of the Session Manager rather than reflect outside knowledge of the topics. This meant that the questions had to elicit basic, navigational decisions. Because the questions were based on actual public-service examples, this was easier to manage. A more difficult consideration was that in order to avoid a ceiling effect, questions had to be chosen that were not too simple. Fortunately, there was a group of questions that were common, but complex, enough to meet both requirements. To make the results more applicable out-

side the existing Session Manager structure, and thus to preserve external validity, underlying design concepts rather than isolated interface problems were targeted. To refine the questionnaire, a sample test was administered to a group of twelve undergraduate students.

Questionnaire

The test questions were drawn from a group of actual information requests reported by public-service staff: being able to distinguish library catalog databases from index databases; selecting appropriate databases from subject categories; locating campus information sources such as directories and class schedules; and accessing the Internet. The questionnaire was divided into three parts. The first part provided information about the subjects to make sure that the person filling out the questionnaire matched the authors' established user profile (undergraduate students) and to determine whether he or she had received any library instruction in the past two years. The questionnaire also asked whether students spoke English as their first language (although this factor was later determined to be insignificant in the overall results). The next part of the questionnaire tested the Welcome screen of the Session Manager, where initial navigation decisions are made. The final part tested the selection of databases from the All Databases screen where secondary database selection occurs.

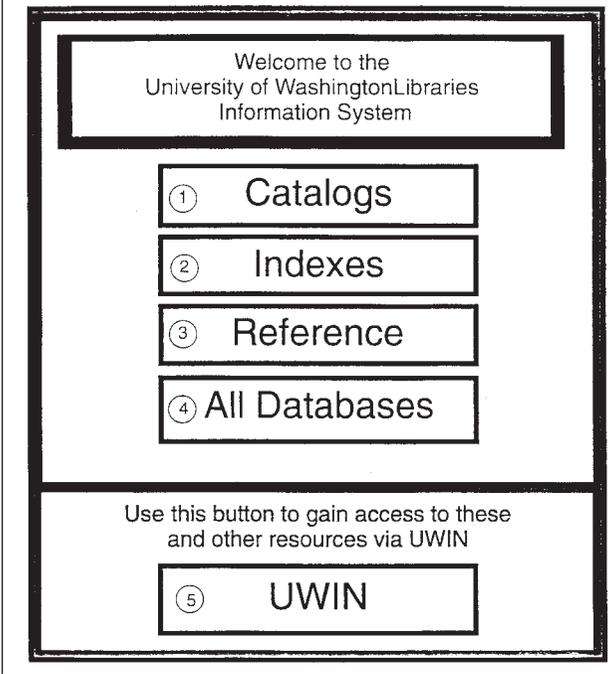


Three variations of the Session Manager introductory and secondary screens were developed to be used with the standard set of questions. Version A duplicated the current Session Manager, version B maintained the current layout but expanded the terminology, and version C maintained the current terminology but changed the layout by grouping resources according to content (see figures 2, 3, and 4).

Administration of the Questionnaire

The questionnaire was administered to 170 undergraduate students. With the helpful approval of two faculty members, large undergraduate social science lecture classes were targeted. As students entered the classrooms, they were randomly assigned version A, B, or C of the questionnaire and were given ten minutes to complete it. The testing took place during

FIGURE 4
Low-fidelity Prototype of Existing Session Manager with Altered Groupings/Layout, Version B



decided ahead of time on the “best” answer and all the “right” answers. The “best” answer was defined as most direct. The “right” answers would all lead to appropriate information but would require more steps; “right” answers were less efficient than “best” answers. “Wrong” answers were defined as choices that would not retrieve information useful for answering the question. In scoring the questionnaire, the “best” and “right” answers were tallied and compared. Students did not select significantly more “right” answers than “best” answers on any version of the questionnaire. All results dis-

class time; no compensation was given to participants. Twenty-five subjects who were outside the established user profile, or who filled out less than one-third of the questionnaire, were eliminated from the study. This created unequal subject cell sizes for the three versions of the questionnaire: A = 55, B = 48, C = 42. Also, because subjects had been randomly assigned, there was an uneven distribution of students who had participated in library instruction. The choice and analysis of the statistics take into account these inequalities. Moreover, a test was run to be sure there was no significant difference between the test results of the two classrooms of students tested.

When scoring student questionnaires, the committee recognized that some questions had more than one correct answer. For every question, the committee

discussed in this article are based on the “right” answer scores.

Results and Discussion

A statistical analysis of variance (ANOVA) was used to determine whether screen image had any effect on the competency test scores. The results showed a main effect for screen image. A statistical (Scheffe) test showed that the difference in scores on screen image testing was significant in the case of both terminology and layout/grouping; students who filled out questionnaires with altered layout or altered terminology scored significantly better than those who used the existing Session Manager.³ There was no significant difference between the scores of the students filling out versions B and C of the questionnaire; layout/grouping and terminology positively affected scores similarly.

The results of the ANOVA also showed a main effect for training. The subjects who had attended a UW Libraries instruction class in the past two years had a significantly higher mean score. These students not only scored higher on the existing Session Manager but also outscored the students without library instruction on the two altered versions. The results of the study upheld the committee's hypothesis that grouping resources and assigning concrete, descriptive labels help undergraduates, especially those with basic library instruction, to make more efficient navigation decisions. Concretely, this means that in order for undergraduates to make efficient choices in a complex information environment, more descriptive text is necessary. In addition, grouping online resources by content rather than by subject provides more effective navigational structure. These results already have been used in designing the new version of the database selection screen for the uniform interface and for current Web development. Because hypertext navigation consists of grouping information, and often leads to design decisions about the length and explicitness of accompanying text, the results of the Session Manager study has helped guide the creation of Web-based tools.

In gathering the data necessary to address the specific problem statement, the committee also hoped to find practical answers to the following questions: Where are students making the most mistakes using the current Session Manager? What effect do screen changes have on initial and secondary navigation choices? What

design elements and terminology choices deserve future investigation? Students consistently made errors on certain questions. For example, only 21.8 percent of the students using the existing Session Manager could find the location of the *New York Times* correctly. Even fewer, 18.4 percent, could determine where to look for a book at the Health Sciences Library. Students also showed a greater ability to make primary, rather than secondary, screen navigation decisions. Although the average score on the first part of the existing Session Manager test was 77 percent, the average score on the second part, dealing with secondary database selection, was only 54 percent. Deciding why these results occurred lies outside the scope of the present research design, but the results point to areas for further examination.

Conclusions and Recommendations

Now that the committee has outlined two basic ways that presentation and description of databases can be improved, and has identified places where undergraduates are making mistakes using the current Session Manager, it has begun to focus on finding out more about why these mistakes occur. A rigorous usability study, using think-aloud protocols, timed-task analysis, and a posttest debriefing will help explain the current results. Sample testing already has revealed interesting patterns of user navigation behavior. In future quantitative and qualitative studies, the authors plan to continue to test and observe user behavior in the evolving online environment.

Notes

1. For more information on the early development of Willow, see Ketchell, Fuller, Freedman, and Lightfoot, "Collaborative Development of a Uniform Graphical Interface," *Proceedings of the Annual Symposium on Computer Applications in Medical Care* (1992): 251-55; and Freedman, "The University of Washington Information Looker-Upper Layered over Windows," *X Resource* 14 (Apr. 1995): 13-31. To view Willow and find out more about its current development, contact:

<http://www.washington.edu/willow/>.

2. Relevant research articles were found by searching the literature of many different disciplines, including library science, technical communication, cognitive psychology, human factors, and computer science. Related bibliographic information can be found in Karen Eliassen, "Research Methodologies: A Complementary Approach in Library and Information Science" (Thesis, University of Washington, 1996).

3. The test scores were examined and compared using Statview statistics software on a Macintosh computer and SPSS statistics software on a PC. The main effects reported here are based on ANOVA results with significance values $p = <.01$ and comparison of means with standard deviations < 2.12 . The Scheffe test was significant at $p = <.05$.