Evaluating the Performance of the Online Public Access Catalog: A Redefinition of Basic Measures

Robert N. Bland

The evaluation of automated library systems is a complex task that generally consists of three components: a functional evaluation, an economic evaluation, and a performance evaluation. The point of a functional evaluation is to determine whether a system provides those features or functions which a library desires. For an online public access catalog (OPAC), these features may range from the ability to do direct Boolean searches with a specified number of operands and operators, to the ability of the system to monitor and provide statistical reports on port activity. An economic evaluation, on the other hand, is an attempt to determine whether a given system is reasonably priced and/or priced within the budget of a library considering purchase. This assessment will normally include an analysis of initial hardware, software, and training costs as well as an analysis of ongoing maintenance and personnel costs. It may also include an analysis of the financial health of the vendor. Finally, a performance evaluation is an attempt to determine whether a system performs or executes its functions within limits negotiated between the vendor and the interested library. These limits will refer to reliability (the amount of "down" time that is acceptable); to capacity (the number of terminals and records supported, and so on); and to performance in the narrower sense (i.e., the speed at which the system runs and the efficiency with which it produces results). Capacity and operational performance are usually evaluated on the basis of a benchmark or acceptance test, in which the system is put through its paces in a configuration and an environment similar or identical to the one in which it is planned to operate.

The library literature is replete with studies and recommendations regarding the functional evaluation of OPACs. Matthews, Hildreth, Fayen,[1] and others have done an admirable job of identifying and explaining the functions which state-of-the-art OPACs should include. This literature has such breadth and depth that even libraries with limited expertise and experience with automation should have little trouble in choosing between alternative OPACs or thoroughly evaluating a single system based on functional characteristics. Through the RFP process and the sharing of information gleaned from contract negotiations, libraries generally have also had considerable information upon which to base economic evaluations. Performance evaluation is another story. Although there is general agreement on standards of reliability inherited from the wider computer industry, there seems to be little formal or informal consensus within the profession regarding the other aspects of performance evaluation. As a consequence, libraries too often find themselves acquiescing to performance evaluations based on performance measures defined by system vendors. These evaluations may not tell a library what it needs to know: that is, whether the complicated and expensive system it is considering or has purchased is really adequate in terms of power and speed for the job intended.

The computer industry has developed many methods for measuring the performance of computer systems. These range from determining the millions of instructions per second executed (MIPS) and the percentage of CPU activity consumed by certain jobs, to determining seek, read, and data transfer times for input-output operations. It is not the business of librarians to be concerned with these measures directly. What is important is how the capacity and power of a computer system translate into productivity for the library application and its intended objectives. The most familiar performance measure associated with automated library systems in general and OPACs in particular is *response time.* Most library-generated specifications for automated systems nowadays contain clauses regarding system response time, usually both for normal

Robert N. Bland is Associate University Librarian for Technical and Automated Services at Ramsey Library, University of North Carolina at Asheville.

and peak load conditions. These terms are defined elsewhere in the system as the number of terminals active, the number of jobs running at a given time, and so on. The mean response time for OPAC searches under normal conditions, the specifications may state, shall not exceed three seconds. Mean response time for OPAC searches under peak load conditions, the specifications may continue, shall not exceed seven seconds. On the surface this may appear precise and effective. But what is response time? And does it have anything to do with the adequacy of the OPAC under consideration as a library catalog? The answer to the first of these questions, I have suggested, is that too often the definition of response time turns out to be what the vendor, not the library, determines it to be. And therefore, the answer to the second question is that response time may have little or nothing to do with the adequacy of the system as a library catalog.

Consider, for example, the common definition of response time as the interval of time between the instant a user presses the return or send key on the terminal and the instant when characters appear on the terminal screen "in response."[2] Let us examine now hypothetical system A. Assume that system A includes an OPAC that functions much like the typical second-generation OPACs found in most automated libraries today. That is, the system may be used with either menus or direct commands; in response to a search query, a summary screen listing matching entries is displayed; the user determines from this summary screen the records s/he wishes to see in detail; and the full record including call number for each title is then displayed at the user's request. Let us assume now that a user of this system wishes to determine the call numbers of all the editions of Samuel Butler's *The Way of All Flesh* owned by the library, and that the system is performing with an average response time as defined above of three seconds. A command search by title will involve a number of discrete steps in this system, each with its own response time wait. The command must be entered, choices must be selected from the summary screen one by one, and the full records then displayed separately. Let us say that a total of four commands are required to display the full records including call numbers for the three editions of *The Way of All Flesh* owned by the library. Thus, leaving aside for the moment the time required for the system to write information to the CRT screen and, of course, the reaction time of the user, a total of twelve seconds is required for the system to complete the user's task.

Consider now hypothetical system B, that supports an OPAC with a rather advanced natural language command interface. Here the user with the same task in mind enters the command: "find all books with title = 'Way of all flesh' and then print call number." The user waits six seconds, and then the call numbers of the three records are displayed upon the screen. Response time for OPAC B is six seconds, twice as bad as system A. Yet it is plain that the user of system B has accomplished his task in half the time of the user of system A.

What is important is how the capacity and power of a computer system translate into productivity for the library application and its intended objectives.

The same point can be made without reference to a hypothetical system with a powerful natural language command processor.[3] Let us assume that system C is identical to system A in terms of its search logic—that for a given search, exactly the same number of discrete steps are required. The operating system of system C, however, is designed and tuned somewhat differently from system A's. Response time as defined above for system C is four seconds. Yet because of perhaps a faster data transfer rate from disk, larger buffers, and less page swapping from memory to disk, system C is able to complete writing to the screen a bibliographic record complete with call number in one second versus two seconds for system A. Thus, with respect to the same example, the total execution time of the task for the two systems, now including both response time and the time to write the screen, is the same: twenty seconds. To the user, interested in results, the systems are identical; but in terms of response time system A is considerably superior to system B.

This definition of response time, then, can hardly be adequate as a basic performance measure for OPACs. It can be too easily manipulated, and without reference to the number of discrete steps involved to complete a task and the amount of time required to write displays, response time measurements can give little information about the system. Response time so defined may be even less adequate as a performance measure to evaluate other system components, where even a larger number of discrete steps may be required to complete a task than is the case with OPAC searching (e.g., adding patrons to the patron data base in the circulation subsystem). Libraries should not participate in benchmark or acceptance tests based on this kind of performance measure, some of which may border on fraud.

Defining response time as the interval of time between the instant when a command is entered and the time when the system is ready to accept another command is certainly an improvement. Under this definition, which some vendors have accepted,[4] response time must include disk seek, read, and data transfer times, and the time to write information to the screen (which may be significant when the system is under heavy load). Yet this still does not yield a true measure of performance because it does not include the number of discrete steps which may be involved in the completion of a task. The performance of a complex, applications-oriented system like an OPAC depends on both the hardware and the software supporting that application. The performance of system A, with a poorly designed user interface requiring six steps to complete a task, may be worse in terms of genuine results than that of system B, which may require only four, even though system A runs on a larger and faster computer.

... we have fixed for too long on the dubious notion of response time as the basis of performance measures.

What we need are measures of performance that gauge the system as a whole with respect to the applications in which we are interested. For a library OPAC, the amount of time taken to complete actual catalog tasks (not computer tasks) should be the basis of these measures. In the computer industry this concept is sometimes referred to as *throughput*, that is, the amount of useful work that the system can perform under specified conditions in a specified amount of time. Alternatively, it may be referred to as *turnaround time*, the amount of time required to complete a particular job or task. For a library OPAC system, throughput is (or should be) the number of catalog searches (on average) that can be *completed* under specified conditions in a specified amount of time. Turnaround time is (or should be) the amount of time (on average) required to *complete* various catalog search tasks.

But how do we measure throughput in an OPAC and when is throughput satisfactory? We cannot simply have users perform random catalog searches, measure completion times, take an average, and then ask if they are happy. There are too many uncontrolled variables to make this kind of test very meaningful, although it is probably more revealing than the response time tests to which we have become accustomed. The system may seem to have performed poorly because some of the users were unfamiliar with it or were poor typists. It may seem to have performed well because too many easy searches were done.

I do not believe that we can expect to learn much about the performance capabilities of complex systems like library OPACs with such unsystematic approaches, any more than we can expect to learn a lot about the performance of automobiles by placing unskilled drivers in them and asking them to drive around. In both cases we may learn something important—that is, how easy the OPACs or autos are to use. But that is a different question from how well they perform.

To test performance—and particularly to compare performance between systems—we need reasonably skilled operators at the controls putting the systems through standard tests under controlled conditions. For OPACs this means librarians or other skilled users performing a variety of searches under operational conditions and carefully recording the task completion or turnaround time for each type of search. These searches should include the common known-item author and title searches, subject searches with and without subdivisions, catalog browsing, and so on. In other words, it should include those kinds of searches which librarians and library users actually perform and which the system should be designed to handle with facility. Moreover, for fair comparisions, the searches should be categorized on the basis of the number of items initially retrieved by the search. We should collect data on the task completion or turnaround time for searches when ten or fewer items are retrieved, when eleven to one hundred items are retrieved, and so on. We will then have a basis for accurately judging how a system will perform in our library *as a library catalog*, and for comparing it with other systems.

But how well should an OPAC perform? Is it reasonable to expect that title searches retrieving fewer than ten titles initially, for example, be completed on average within ten seconds? thirty seconds? two minutes? It is somewhat curious that in all the library literature on OPACs little or no data regarding this question appears, and this lack, I suspect, is because we have fixed for too long on the dubious notion of response time as the basis of performance measures. To be sure, many surveys designed to determine the satisfaction of library patrons using OPACs have been done, and they are no doubt important and useful in their own right. Yet without quantification under controlled conditions of the turnaround time of the various activities being performed, we cannot use these studies as a basis for the development of objective standards of OPAC performance.

In an attempt to establish some baseline figures for catalog activity turnaround times in the online catalog, a study was made by the author, during the summer, of six different OPACs (representing three different systems) installed and operational in North Carolina academic libraries. Two of these OPACs were large systems, holding more than 500,000 bibliographic records; two were middle-sized systems, with between 200,000 and 300,000 bibliographic records; and two were small systems, with fewer than 150,000 bibliographic records. Five queries were made of each OPAC on each trial over a two-week period in order to determine baseline figures for the following types of catalog searches:

- Search Type #1 An author search retrieving fewer than ten records, one of which (the first) is selected for full display;
- Search Type #2 A title search retrieving fewer than ten records, one of which (the last) is selected for full display;
- Search Type #3 A subject search retrieving fewer than ten entries, one of which (the first) is selected for display;
- Search Type #4 A subject search retrieving between ten and fifty entries, the first three of which are displayed as full records;
- Search Type #5 An author search retrieving between ten and fifty records, two of which (the first and last of the list) are selected for full display.

This list is, I think, fairly representative of the types of searches actual library users do (or attempt to do) in our catalogs, although it is certainly not complete. To maintain consistency in the comparisons, the actual searches done in the test were pre-selected by examining each catalog and identifying a search in each category which would retrieve approximately the same number of records as retrieved in the other catalogs. The searches were all entered as direct commands, even where menu selections were available. Turnaround time was measured as the interval between the instant when typing of the command began to the instant the last full record of the set was displayed, complete with call number. Thus, turnaround time for each type of search included the author's keying time (I am a fairly decent typist—fifty words per minute), the time for the OPAC program to decipher the command, perform disk reads to retrieve the necessary information, display the information on the screen, and wait for the next command. The author's reaction time was included, but in this case, at least, that time was minimal because it was predetermined which records from the summary screen would be further selected for full display.

The results of the test for each OPAC appear in Table 1. The figures in the bottom row indicate average search times for all the OPACs for the type of search indicated at the top of the column.

The 1989 Book Week poster has been created for the Children's Book Council by Caldecott Medalist Richard Egielski. National Book Week will be observed for the 70th year November 13-19 in 1989. Egielski's full-color 17" x 22" poster costs $6.50. It is also available with many other items at a substantial savings in the Council's Book Week Bargain Kit. Send a 25¢ stamped, self-addressed envelope to CBC (P.O. Box 706, New York, NY 10276-0706) for "Book Week Brochure" for details.

TABLE 1.
Online Catalog Searches

System	Number of Bibliographic Records	(Mean Turnaround Times) Search Types #1	#2	#3	#4	#5	ALL
OPAC 1	500,000+	21.8	19.8	21.6	54.9	46.0	32.8
OPAC 2	500,000+	21.7	18.5	19.9	70.6	69.2	40.0
OPAC 3	200,000-400,000	18.0	12.7	12.0	41.9	61.5	29.2
OPAC 4	200,000-400,000	25.5	23.4	39.9	79.2	75.7	48.7
OPAC 5	50,000-150,000	31.4	18.0	29.4	61.4	84.6	44.9
OPAC 6	50,000-150,000	23.3	17.6	18.9	54.5	48.8	32.6
Column Means:		23.7	18.5	23.9	60.9	64.4	38.2

A GLM (general linear models) procedure used with the SAS statistical package to analyze variance showed statistically significant differences among the OPACs for all five search types: $p < .05$. However, Tukey's Studentized Range Test, used to test overall performance in which each OPAC was compared against each of the others on all search types, showed statistically significant differences at the .05 significance level only between OPACs 3 and 4. The differences among all the other OPACs were leveled out somewhat by one system performing better in one search type and another better in another search type.

Intuitively, the search times for the OPACs generally seem quite good. It should be remembered, however, that the searches were done with the systems under relatively light loads during the summer when academic library use is at low ebb. It will be interesting to see how these figures change with the systems under heavy user load, as we may expect them to be during the peak periods of fall semester. Based on previous although undocumented experience, we can expect significant degradations in turnaround time for some of these systems.

It is interesting, too, to compare the results in Table 1 with Table 2, which show average turnaround times for the first three searches in manual catalogs—in one case, a card catalog for a middle-sized library and, in the second, a computer output microform (COM) catalog for a small library.[5] A minimum standard for online catalogs in the marketplace today might be that known-item or single-entry searches like those described in the first three test searches should be at least as efficient as manual searches in catalogs of similar size even when the system is under heavy load. Manual catalogs often actually perform better in terms of turnaround time in searches of the type described in (4) and (5), where the user's task is to identify a number of titles under a single heading, resulting from the ease with which one can move from one full record to the next in a card or COM catalog once the appropriate position in the index has been reached. Analysis of the differences in turnaround times for various catalog tasks performed in automated and manual systems can help to identify types of searches which online systems perform well and those which they perform less well. It can also help system designers to search for ways in which to overcome or circumvent the inherent physical limitations of the CRT screen.

TABLE 2.
Manual Catalog Searches

System	Number of Bibliographic Records	(Mean Turnaround Times) Search Types #1	#2	#3
Cards	200,000-400,000	26.9	25.9	34.4
COM	150,000	33.2	34.0	36.0

At any rate, such a preliminary investigation cannot, of course, establish standards of itself, even for systems under light load. The point of the investigation is to show how meaningful performance measures for OPACs (or any other library system, for that matter) can be developed and used to evaluate performance in a manner which allows us to compare systems in a reasonably objective manner with respect to actual library applications, and even to compare automated systems against manual systems. Much further work will be required before true standards for OPAC performance can be developed, but there can be no doubt that standards are needed. An online catalog represents a vast investment for a library, in terms of both financial and personnel resources. We are past the time when decisions of this magnitude can properly be made on the basis of word-of-mouth, subjective user opinions, or poorly defined performance measures. Standards will also help library automation vendors to understand library needs and to develop systems that truly meet library functional and performance requirements. The current environment, lacking as it does a consensus regarding performance measures and objective performance standards, virtually invites competitive bid situations which result in libraries' purchasing systems that are underpowered or otherwise inadequate to library performance needs. With improved measures of performance in place and solid empirical data upon which to base expectations, libraries and system vendors alike should be better prepared to avoid costly mistakes in the future.

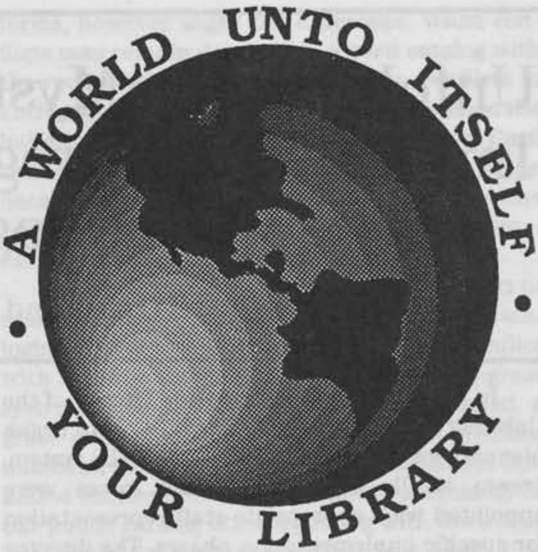
References

1. Joseph R. Matthews. *Public Access to Online Catalogs* (Weston, Conn.: Online, Inc., 1982). Charles R. Hildreth. *Online Public Access Catalogs* (Dublin, Ohio: OCLC, Inc., 1982). Emily Gallup Fayen. *The Online Catalog* (White Plains, N.Y.: Knowledge Industry, 1983).

2. Matthews, p. 53.

3. For those unfamiliar with natural language command processors, the hypothetical command above is by no means science fiction. Once a data base for books has been defined with call number as a field, this exact command could be used in Digital Equipment Corporation's *Datatrieve* report and query system to perform the task described.

Barry Moser has created a poster on *1990 International Literacy Year* for the Children's Book Council. The poster, measuring 24" x 32½", is in six colors. It costs $15.00 and is shipped rolled in a mailing tube. Send a 25¢ stamped, self-addressed envelope to CBC (P.O. Box 706, New York, NY 10276-0706) for *Current Materials Brochure* for details.



Library Administration and Management Section Formed

A new section of NCLA named "Library Administration and Management Section" (LAMS) has been approved by the Executive Board of NCLA. LAMS will hold its first program and organization meeting at the NCLA conference in October in Charlotte. Patterned (to some degree) after LAMA, a division of ALA, the mission of LAMS of NCLA will be to provide an organizational framework for improving the practice of administration in libraries and for identifying and fostering administrative skills. The section will meet its responsibility by aiding the professional development of personnel interested in administration and management and by planning and developing programs, study, and research in library administration and management problems.

NCLA members who are interested in administration and management are encouraged to designate a preference for this section (LAMS) at the time of payment of biennial associational dues. NCLA members may join LAMS any time during the year, however, by notifying the treasurer and paying section dues.

All attending the NCLA conference are cordially invited to attend the program and organizational meeting of LAMS on Thursday, October 12, from 9 to 10:30 AM.

For additional information, one may contact LAMS Steering Committee Chair, Miss Nancy Ray, Director, Southern Pines Public Library, 180 S.W. Broad St., Southern Pines 28387, Phone: 919-692-8235.