Developing Meaningful Measures for Electronic Reference Services: New Technology, Old Questions

Eric Novotny
History Subject Specialist, Pennsylvania State University
State College, Pennsylvania, USA

Abstract

Throughout its history, the library profession has been bedeviled by the question of how to measure the value and usefulness of our services. In an encounter as complex as the reference interaction, how can we tell if we are serving our users as effectively as we possibly can? The effects of the Internet and information technology have only added to the chaos. As many libraries worry about losing their users to commercial information services, assessment has taken on a new urgency.

In this presentation I will examine the current reference service environment and how this context affects the climate for assessment. I will also take a step back in time to review existing evaluative methods and criteria. While some of the issues are new, the basic questions of effectiveness and value remain the same. It should not be necessary to reinvent the assessment wheel with every technological advance. I will conclude with a few concrete examples of qualitative criteria that can be adapted to online reference services.

Current Reference Service Environment

Assessment does not take place in a vacuum. The context we operate in shapes the questions we seek to answer. An environmental issue that is having a significant impact on reference services is the widespread phenomenon of the vanishing reference questions. My own institution recorded a dramatic decline in reference desk questions, falling from 504,321 inquiries in 1995 to 352,647 in 1999. In part this can be attributed to local circumstances (a massive construction project in the library), but a similar trend has been reported throughout American research libraries. The median reference statistics reported by ARL member libraries dropped from 162,336 in 1997 to 129,482 in 1999.[1]

The reasons for the decline in reference statistics remain the topic of much speculation and discussion. It can be argued that the decline is cause for celebration rather than alarm. Fewer questions at the reference desk may reflect the increased ability of our users to find the information they need without our assistance. If this is truly the case then we should be patting ourselves on the backs for our success in organizing information so effectively that our users no longer need us!

Others have pointed to problems with our data collection methods as the explanation for declining reference statistics. As the library liaison to the history department, I spend significant amounts of time responding to e-mail and phone queries sent directly to me. These interactions are rarely recorded. There is some evidence that mine is not a unique experience. When Ohio State University began recording off-desk reference interactions, they discovered that those activities consumed almost as much of their time as their scheduled desk hours.[2] While the Ohio State librarians did not answer as many questions in their offices as they did at public service points, they generally spent more time with each off-desk question.

A satisfactory explanation for why our reference statistics are going down almost certainly involves a combination of many factors. All of the uncertainty has drawn attention to the shortcomings of relying solely on quantitative data for assessing reference services. Lacking other forms of information, the library profession is reduced to educated guesswork on this very important question. Ultimately no one really knows how to interpret the decline in reference statistics. What is clear is that we need to develop and implement methods to generate better data if we are to advance beyond speculation.

A topic that has generated almost as much discussion as the drop-off in in-library activity is the mysterious and persistent failure of most electronic reference services to capture the imagination of our users. In the early days of e-mail reference, libraries adopted stringent policies that were designed to limit usage. Anticipating a horde of computer-savvy users from all corners of the globe, many libraries placed restrictions on who could use the service, when, and how. Unaffiliated users and others such as genealogists were discouraged from making use of e-mail reference services. Unfortunately, it appears we have become too successful at deterring use. The members of the Association of Research Libraries reported receiving an average of 67 questions per month (or a little more than 2 per day) in an October 1999 SPEC Kit survey. The majority of these libraries indicated that their usage was increasing, but only at a moderate pace.[3] Some library chat sites have reported significantly larger numbers, but they typically pale in comparison to commercial services like AskJeeves, which receives upwards of 100 million inquiries a month.[4]

Intuitively many librarians remain convinced that remote users are floundering and need our help. This may be the case, but the numbers indicate that they are not turning to libraries in their hour of need. More research is needed to uncover why our users have not taken full advantage of our electronic service offerings. This requires talking to those who do not avail themselves of our services to find out what factors contribute to non-use. Once the data has been collected it then needs to be disseminated. Reports of successes (and failures) can contribute to a shared knowledge base informing local practice.

A third factor in the reference environment that impacts assessment is the "invisible" remote user. Public service librarians at a reference desk were often able to rely on informal methods of feedback. A patron's smile or a momentary grimace would communicate attitudes about library resources and services. Body language and offhand comments regarding the "ugly" stacks or the "always broken" copies were visual and verbal cues as to how are patrons perceived the library. Working the reference desk we would frequently receive instant feedback regarding user satisfaction. A large percentage of patrons walking away smiling was a pretty good indicator of an effective service (not to mention being personally gratifying). Frequently we lose such visual cues when communicating with our patrons electronically. The absence of this type of feedback makes it harder to get a sense of how the encounter is progressing, and makes it more important to use formal evaluative methods to ascertain our effectiveness.

In spite of vanishing users and declining desk statistics, most librarians do not appear to be leading less hectic professional lives. Librarians have eagerly taken on new responsibilities, while at the same time maintaining existing services. This is only natural. It is often easiest to delay hard decisions by offering a new service without eliminating or reducing existing services. While this may work for a while, it is not a recipe for long-term job satisfaction. It is not possible to work longer and harder indefinitely.

Inevitably choices need to be made. We can offer many services that are of value to our users. We cannot, however, provide every potentially useful service our users might want. In an environment of fixed resources, libraries cannot be all things to all people. There is a need to prioritize amongst the many service options available. A consistent, coherent program of evaluation can make choosing amongst the available options easier by providing sound data on which to base decisions.

This is the environment we are operating in. We have lots of questions and no perfect mechanism for obtaining answers. The good news is that there is no shortage of approaches for getting the information we need. In fact there are too many to discuss in any great detail here. It is, however, useful to briefly discuss the chief methodologies. Those looking for a more in-depth discussion are referred to the recent publications by Hernon & Whitman, Bertot, McClure & Ryan, and Whitlatch.[5]

Assessment Methodologies

Probably the most widely used evaluative tool in libraries is personal observation and experience. Ask any veteran practitioner about their service and they can usually provide an estimate of the level of activity, and their perspective on the positive and negative aspects of the service. This information is often supplemented by gathering what are (hopefully) more precise records about the service. Collecting available data is an important component of assessment that will be discussed in more detail later.

An electronic variant on personal observation is the transaction log. Originally used to track user behavior in online catalogs and databases, transaction logs are standard features of most chat reference utilities. They have tremendous potential in service assessment, as they can automatically create a record of the reference encounter. This has tremendous potential not just for reference evaluation, but also for training and professional development. The main concern with transaction log tools is privacy. The institution will need to develop procedures to reassure staff and patrons who may be worried about "Big Brother" observing their reference encounters. While this issue should not be downplayed, neither should it unduly interfere with efforts to improve reference services. Stripped of identifying personal information, the recorded encounters provide opportunities to discuss research strategies and resources in a spirit of cooperative improvement.

Surveys and focus groups are widely used techniques and are well covered in numerous guides to research.[6] Briefly, surveys are best used if you want a standardized set of responses from a large group of people. Interviews or focus groups (group interviews) allow for more in-depth probing of feelings and attitudes. Often these are effectively utilized in tandem. One approach to evaluating a new chat reference service might be to conduct a survey of users of the service. This could then be supplemented with information from a focus group of non-users to ascertain their thoughts about the service.

I would like to focus a bit more attention on two research methods that are less frequently used in reference assessment. The first is called contingent valuation. This method of analysis utilizes the concept of willingness to pay. This technique is described in a study conducted at Virginia Commonwealth University. Patrons who agreed to participate in the study were given information on the cost of selected campus services (e.g. the art museum, parking, etc.). They were then asked to estimate how much they would be willing to pay to maintain existing reference hours, and how much they would pay for expanded service hours.[7] Two key findings of this study were; 1) users placed a high value on reference service, even if they did not plan to use it themselves, and 2) the value assigned by users for current services was about 3.5

times the estimated cost of providing the service. Similar studies focusing on electronic reference could help to establish the relative value of such programs. It may be that our users, who are largely ignoring library e-reference services, still see great value in having the option of sending a remote query at 2:00 a.m.

A relatively recent research method that is becoming increasingly popular is the usability study. This approach is typically employed to assess the ease of navigation of an institution's web site. Users are given a set of tasks and their performance is recorded and analyzed to see how well they were able to perform the assigned tasks. Library studies using this methodology have asked users to locate online research resources such as the online catalog, or a listing of e-journals. While not specifically designed to measure services, I believe that this approach can provide useful information for the evaluation of electronic reference services, particularly the question of why so many receive relatively little use. A recent usability study at the University of Newfoundland found that seventy six percent of the participants could correctly identify where to go for help from the library's front screen.[8] This discovery can be interpreted a number of ways, depending on one's perspective. I would argue that there is cause for concern if almost a quarter of our users are unable to even locate our electronic reference service. This would automatically exclude one in four potential questioners! Also of concern is the fact that none of the users in the study reported having ever used the service before. Findings like these raise questions about the accessibility of our e-reference services. We need more such research to determine if we are reaching our users effectively.

Quantitative Methods

A decision that is closely linked to the choice of methodologies is the question of what to collect. What you decide to measure will help determine the approach you use to gather information. Probably the most common approach is gathering statistics from existing records. This technique is hardly unique to e-reference, but what has changed in the digital environment is the amount of information that can be preserved for analysis. My previous institution assigned a staff member the duty of manually recording data from all of the e-mail inquiries we received. Now, many chat reference programs can be configured to automatically record such information. As we gain in our ability to gather data it is useful to consider what we want to know about our electronic reference transactions. I have included below a few examples of statistics that might be collected for your consideration. For a good example of data collection and analysis of a chat reference service see the report of the Ready for Reference project.[9]

- Number of users/Number of sessions
- Number of questions asked
- Category of user/User affiliation (undergraduate, faculty, general public)
- Length of reference encounter
- Day/Time of question
- Response time
- Number of repeat users

The problem with quantitative data is not so much the collection as the interpretation. With chat reference logs you can discover with a high degree of precision that you averaged six users per hour, that the average transaction took 14 minutes, and that 82% of your users were undergraduates. The question then becomes, so what? Information in the abstract is of limited usefulness. What is needed is context. How does e-mail or chat reference compare with in-person reference? If there are differences are they important and if they are, what can be done to address them?

Of course, this type of comparative analysis assumes that you have similar information about related services. The first survey I did was an examination of the OCLC Firstsearch service. Our evaluation team eagerly gathered as much data as we could. We also polled users and discovered that x% of respondents were satisfied with Firstsearch after using it. Unfortunately we had no comparable data for other services, so were baffled as to how to interpret this information. We had no idea if the recorded level of satisfaction was higher, lower, or about the same as that expressed by users of other products. A comprehensive program of continuing assessment can fill in these gaps.

Besides facilitating comparisons between services, quantitative data can also be used to compare the service with itself. Data collected over time can be especially useful in documenting user attitudes before and after changes in services. Benchmarking against prior performance is also a good way to continually raise the bar. If eighty percent of your e-reference users are satisfied this year, you can look for ways to ensure that eighty five percent are satisfied next year.

Finally, it is human nature to want to use numbers when comparing ourselves to others. While this needs to be done with some caution, it does serve the purpose of situating the institutional data in a broader context. Some comparative data is available for electronic reference services. The members of the Association of Research Libraries (ARL) were polled in a SPEC Kit survey on Electronic reference service. Responses indicated that the average level of activity for ARL libraries was 67 questions per month, excluding outliers such as the Library of Congress and the National Library of Medicine.[10] The numbers are from 1999, but the libraries surveyed did not anticipate a sudden rise in their e-reference activity. While seventy five percent reported a gradual increase in usage, only one respondent indicated that the number of questions were increasingly rapidly.

The SPEC Kit responses apply primarily to e-mail reference, as most libraries did not offer chat reference at the time the survey was conducted. What then are the usage statistics for chat reference? Comprehensive data is not currently available, but some institutions have chosen to publicize their usage statistics. The Ready for Reference project, a consortium of Illinois libraries reported 474 sessions during the period February-April.[11] A service offered by the Business library of the University of Chicago received an average of two questions/day.[12] In the absence of comprehensively collected information those desiring use statistics will most likely need to contact libraries offering similar services. This task is made easier by the existence of online registries that link to libraries offering e-mail and chat reference.[13]

Qualitative Methods

While statistics can be useful, numbers by themselves rarely tell the whole story. It is possible for a widely used service to be quite unpopular if there is no acceptable substitute (e.g. the local OPAC). Similarly it is possible for a less widespread service to be highly valued by those few who use it. Ultimately we owe it to our users to determine not just if the situation is satisfactory, but also if we are providing the best possible mix of resources and services. This requires that we look beyond the number of users and consider qualitative factors such as satisfaction, usefulness, and efficiency. Ultimately cost must also be a consideration. Money spent on one service invariably limits the available funds for other, potentially more useful services.

The library profession does not lack criteria with which to measure reference quality. In fact there are so many that it is not possible to discuss the entire range of proposed measures. A few qualitative measures that I believe could be applied to assessing electronic reference service include:

- Accuracy of Response
- Approachability
- Cost
- Effectiveness in meeting needs
- Efficiency
- Impact
- Reliability
- Response Time
- User Satisfaction
- Usefulness/Value
- Willingness to Return

In general the criteria listed above fall into one of two categories. Some are outcome measures that seek to capture the results of the transaction. (Was the response accurate, was the user's need effectively met?) Other criteria are input measures. They look at factors that contribute to the success or failure of the enterprise. Reliability and response time are examples of input measures. Having data on these factors can help us to interpret why we achieved the outcomes that we did.

Most of the measures I have listed are not new. Approachability and accessibility are two criteria that have been around for some time. They appear as factors in reference desk success in the RUSA Behavioral Guidelines for Reference Service Providers.[14] Although these measures were originally created to evaluate reference services provided in person at a reference desk, I believe that it does not take much creativity to adapt them to the online environment. The examples below show how a program of assessing electronic reference services might be implemented. The percentages used are arbitrary and are included merely for demonstration purposes. We need more data before we can develop true benchmarks. Once collected, the qualitative measures can be usefully combined with quantitative measures such as number of users and response time, to create a more complete picture of an e-reference service.

Applying Qualitative Measures to Digital Reference Services: A Few Examples:

Measure: Accessibility
Methodology: User survey, Focus group, or Usability study
Benchmark: 95% of population can correctly identify and access online reference services
Measure: Approachability
Methodology: Observation, Surveys, Focus groups, Usability study
Benchmark: 95% of population rate e-reference site as welcoming and friendly
Measure: Accuracy of Response
Methodology: Transaction analysis, Survey
Benchmark: 80% of answers are accurate (user or peer librarian assessment)
Measure: Willingness to Return
Methodology: Transaction logs, Survey, Focus group
Benchmark: 90% of users express a willingness to use the service again as needed

Electronic Reference Assessment: Where do we go from here?

I have presented some approaches to the assessment of electronic reference services. There is no shortage of techniques for those wishing to assess their reference programs. What the profession does lack are standards and benchmarks by which to judge performance. There are only a few studies presenting qualitative data on the use of electronic reference services.[15] This is not especially surprising given the relative newness of e-reference services. What is more alarming is the profession's inability to reach any consensus on quality standards for more traditional reference activities. Based on this it appears that it will be some time before the library community is able to develop quality standards for digital reference

There are some things you can do as a researcher to hasten progress. Most importantly, don't conduct your research on an intellectual island. As you begin designing your project consider ways that you can expand upon or replicate existing studies. We need more comparable studies to create a critical mass of evidence. Towards this end Dr. Matthew Saxton is creating a web database of reference research studies. Researchers will be able to use the database to identify studies by data collection method and outcomes measured.[16]

While the efforts of individual researchers are essential, the ultimate goal should be a national assessment program in which libraries collect quantitative and qualitative data in a standard fashion. This would reveal broader trends and issues affecting the entire library community. In the United States the feasibility of a national assessment program is being investigated by Charles McClure and David R. Lankes.[17] Their project seeks to gather information on what the library profession should be measuring in the electronic service environment. I urge everyone to take a look at this site and consider how you can contribute to this effort.

While librarians have not always embraced assessment, the future is not without promise. Questions about reference effectiveness have been discussed for decades, but the success of online competitors has given new urgency to the debate. I have suggested some approaches that we can take as a profession to get the information we need. I look forward to continuing the discussion as we collectively explore these issues.

ENDNOTES

1. Association of Research Libraries, "ARL Statistics," <http://www.arl.org/stats/arlstat/index.html> Accessed 8-31-2001.
2. N. Courtney, "Refocusing Reference Statistics: Redefining Statistics to Measure Time Instead of Quantity". Presentation given at the American Library Association MidWinter Discussion Forum, January 14, 2001, co-sponsored by the RUSA MOUSS Evaluation of Reference and User Services committee and the Reference Services in Large Research Libraries Discussion Group.
3. Association of Research Libraries (1999), SPEC Kit #251: *Electronic Reference Service,* Washington: DC., p. 7.
4. The Ask Jeeves 2000 annual report indicates more than 12 billion queries last year. *AskJeeves Inc. Annual Report* <http://www.irconnect.com/askj/files/askjeeves_ar_00.pdf> Accessed 8-31-2001.
5. Bertot, Juan Carlo, Charles McClure and Joe Ryan (2001), "*Statistics and Performance Measures for Public Library Networked Services*," Chicago: American Library Association. Hernon, Peter & John R. Whitman (2001), "*Delivering Satisfaction and Service Quality: A Customer based Approach for Libraries*," Chicago: American Library Association. Whitlatch, Jo Bell (2000), "*Evaluating Reference Services: A Practical Guide*," Chicago: American Library Association.
6. See for example, Whitlatch, Jo Bell (2000), "*Evaluating Reference Services: A Practical Guide*," Chicago: American Library Association.
7. Harless, David W. & Frank R. Allen, "Using the Contingent Valuation Method to Measure Patron Benefits of Reference Desk Service in an Academic library at Virginia Commonwealth University," *College & Research Libraries* 60, no.1 (Jan. '99):56-69.
8. McGillis, Louise & Elaine G. Torns, "Usability of the Academic Library Web Site: Implications for Design," *College & Research Libraries* 62, no.4 (July 2001): 355-368.
9. Sloan, Bernie, "Ready for Reference: Academic Libraries Offer Live Web-based Reference," <http://www.lis.uiuc.edu/~b-sloan/ready4ref.htm> Accessed 8-31-2001.
10. Association of Research Libraries (1999), *SPEC Kit #251: Electronic Reference Service*, Washington: DC., p. 7.
11. Sloan, Bernie, "Ready for Reference: Academic Libraries Offer Live Web-based Reference," <http://www.lis.uiuc.edu/~b-sloan/ready4ref.htm>, Accessed 8-31-2001.
12. University of Chicago, Business and Economics Resource Center, "Ask a Librarian Live," <http://www.lib.uchicago.edu/e/busecon/aboutask.html> Accessed 8-31-2001.
13. Sloan, Bernie, "Digital Reference Services: A Bibliography," <http://www.lis.uiuc.edu/~b-sloan/digiref.htm> Accessed 8-31-2001. "LiveRef: A Registry of Real-Time Digital Reference Services," <http://www.public.iastate.edu/~CYBERSTACKS/LiveRef.htm> Accessed 8-31-2001. "Elite Project," <http://www.le.ac.uk/li/distance/eliteproject/index.htm> Accessed 8-31-2001.

14. American Library Association, Reference and User Services Association, "Guidelines for Behavioral Performance of Reference and Information Service Professionals," <http://www.ala.org/rusa/stnd_behavior.html> Accessed 8-31-2001.

15. Among the formal evaluations conducted are: Lara Bushallow-Wilbur Gemma DeViney, and Fritz Whitcomb, "Electronic Mail Reference Service: A Study," RQ 35 (Spring 1996): 359-371; Ann Bristow, "Academic Reference Service over Electronic Mail: An Update," *College and Research Libraries News* 7 (1995):460; Julie M. Still and Frank Campbell, "Librarian in a Box: the Use of Electronic Mail for Reference," *Reference Services Review* 21, no.1 (1993): 15-18; Beth A. Garnsey and Ronald R. Powell, "Electronic Mail Reference Services in the Public Library," *Reference & User Services Quarterly* 39, no.3 (Spring 2000): 245-252.

16. "Reference Service Evaluation Research," <http://www.ischool.washington.edu/msaxton/reftest/RefTestHome.cfm> Accessed 8-31-2001.

17. McClure, Charles R., & David R.Lankes, "Assessing Quality in Digital Reference," <http://quartz.syr.edu/quality> Accessed 8-31-2001.