

Evaluation of Online Reference Services

by Jeffrey Pomerantz

Virtual Reference Services

Evaluation has always been a critical component of managing an online reference service; indeed, it is a critical component of managing any reference service or even any service, period. Reference, and particularly online reference, is highly resource-intensive work, both of librarians' time and of library materials. Evaluation is the means by which it can be determined if those resources are being used effectively.

Evaluation efforts are essential for reference services for several reasons. Perhaps most importantly, evaluation provides the library administration and the reference service with information about the service itself – how well the service is meeting its intended goals, objectives and outcomes; the degree to which the service is meeting user needs; and whether resources being committed to the service are producing the desired results. In addition, evaluation data provide a basis for the reference service to report and communicate to the broader library, user and political communities about the service. If there is no knowledge of the existing strengths and deficiencies of the service, the services cannot be improved, nor can any worthwhile communication about the service take place with interested stakeholders.

Evaluation data are necessary to assist decision makers in managing a reference service, but they are not sufficient in and of themselves. All evaluation takes place in a political context in which different stakeholder groups (librarians and library administrators, users, state and local government officials, funding sources and so forth) have different and sometimes competing expectations of what a project should be doing and what the results should be.

Jeffrey Pomerantz is an assistant professor in the School of Information and Library Science at the University of North Carolina at Chapel Hill. He may be reached at pomerantz@unc.edu.

Despite the development and implementation of a variety of evaluation measures, different stakeholder groups may interpret evaluation data differently. The data that results from evaluation efforts provide baseline information that can inform decision makers as they discuss the goals and activities of the service. Evaluation data is especially important these days, as more and more libraries are providing online reference services, but the current tight budgetary climate requires libraries to present evidence justifying the value of the services they offer. There have also been some recent cases of chat-based reference services that have been discontinued, and if a library is trying to provide such a service it is important to understand why others have failed.

Any library evaluation can take one or both of two perspectives: that of the library itself and that of the library user. An evaluation from the perspective of the online reference service itself will ask questions concerned with the efficiency of the operations of the service, such as the volume of questions handled per unit of time and the speed with which questions were answered. An evaluation from the perspective of the library user will ask questions concerned with the effectiveness of the output of the service, such as the user's satisfaction with the information provided and the interaction with the librarian. Both of these perspectives are important, and over the long term a service should conduct evaluations from both. The methods that are used to conduct evaluations from these two perspectives are different, however, and so it is often simpler for any single evaluation to take one perspective or the other.

The Perspective of the Library Itself

The collection of statistics is a long-standing tradition in desk reference, but often these statistics are a very thin representation of the reference interaction: how many questions asked per shift, for example, and sometimes the topics of those questions. This lack of detail is due to the difficulty of

capturing transcripts of reference interactions at the desk and the privacy concerns it would raise if one did.

In online reference, on the other hand, the transcript of the reference interaction is captured automatically, as are a range of statistics and other data: email and instant messaging (IM) clients capture a timestamp for when a message is sent and record the librarian's and user's usernames, for example. Commercial reference management software such as QuestionPoint captures more than that, including such data as the duration of a user's queue time before connecting with a librarian. Web server logs may be analyzed to collect additional data, such as the referring URL from which users come to the reference service.

This data enables the service to describe the operation of the service and to identify trends in the use of the service. For example, capturing the timestamps of interactions enables the service to determine the volume of questions received and answers provided by time of day, day of the week and over the course of weeks. Many online reference services find that their volume of questions rises and falls with the academic calendar – and interestingly, this variation is true of services both in academic and public libraries. In IM services, capturing timestamps also enables the service to determine the duration of sessions; many services report average session lengths of approximately 15 minutes. Capturing users' usernames enables analysis of the balance of first-time versus repeat users, either or both of which may be an important metric of success for a service. Capturing the librarian's username enables services offered by a consortium of libraries to identify which libraries' staff are answering more or fewer questions, which can inform the management of workload across the consortium. Capturing the referring URL may enable the service to identify how users find out about the service, which may inform marketing and outreach efforts.

Making use of transcripts of reference interactions for evaluation can be considerably more difficult than making use of descriptive data. Transcripts are in natural language, and while tools for automatically analyzing natural language exist, they are not often used in reference evaluation. Analysis of transcripts is therefore usually performed manually, using methods of content analysis.

While content analysis of transcripts can be quite time-consuming, it is important because it opens reference interactions up for a range of evaluation metrics. The accuracy and completeness of the answer provided can be identified, for example, as can the librarian's adherence to RUSA guidelines and local policies for quality reference service. The conduct of the reference interaction can be analyzed, for example for the use of interviewing and negotiation techniques. Instances of instruction offered by the librarian can be identified in the transcript. A range of measures of user satisfaction can be identified, from the user's comments on the information provided to expressions of thanks.

The trick to conducting content analysis on reference transcripts, however, is rigorously defining what constitutes an accurate and complete answer or an instance of instruction or an expression of thanks. Indeed, this accurate definition is the trick to conducting content analysis on any type of content. For the analysis to be reliable, the scopes of the categories used to code content must be clear.

The privacy concerns in capturing entire transcripts of reference interactions are the same in online reference as at the desk. Many online reference services state the policy that transcripts are kept for some period of time for the purposes of service evaluation, but few state what data about the user may be captured in the transcript or along with it. Of course the user is usually free to provide false data to the reference service, since most services have no mechanism to validate users' personal information. The one piece of information that the user cannot falsify, however, is the question itself, and there may be situations in which the user wishes to keep even this private. The reference service therefore needs to develop policies concerning their use of collected data and the conditions under which users can request that data not be collected. These policies must also be easily available, so that the user can decide whether or not to submit a question to the service.

Some research has also been conducted on methods for removing personally identifying information from reference transcripts, including the application of Health Insurance Portability and Accountability Act (HIPAA) guidelines, though much work still needs to be done in this area.

One last important area for evaluation from the perspective of the library itself is analysis of the cost to the library of providing a reference service. The current tight budgetary climate for libraries has given rise to a renewed interest in cost and return-on-investment studies. Such evaluations can be powerful arguments when addressed to government or other funding agencies, which are increasingly demanding evidence of financial sustainability.

Cost studies are notoriously difficult in libraries, however, because library budgets are often developed such that identification of specific cost drivers is difficult. For example, what percentage of the costs of the library's reference collection and database subscriptions should be attributed to the reference service? What is the cost per answer of reference questions? These questions are unanswerable given most libraries' budgets. A relatively new method for measuring costs is gaining popularity in libraries, however, which may help remedy this failure.

Activity-based costing (ABC) dictates that the services provided be identified first, and then costs can be allocated to those services. This approach is quite different from much current library budgeting, in which costs are allocated by department, even when services are provided across departments. Cost analyses cannot stand on their own, however. They must be combined with other measures of reference service performance. It is all very well for an evaluation to show that a reference service is operating within some budgetary limit, but it requires other measures to argue that offering the service is worthwhile in the first place.

The Perspective of the Library User

A reference service could not exist without users, so users' perceptions of the service are a critical part of any evaluation. Many of the old chestnuts from desk reference evaluation apply equally well for online reference evaluation, such as the completeness of the answer; the user's satisfaction with the librarian's helpfulness, politeness and interest; or his or her willingness to return. There are in fact a wide range of user satisfaction measures, each of which captures a different aspect of satisfaction or satisfaction with a different aspect of the service. It is therefore important

for evaluators to carefully consider which aspects of satisfaction are important to the service's stakeholders. Further, the online environment changes the interaction between the librarian and the user and the user's perception of that interaction, so evaluation metrics of interpersonal interaction must be changed accordingly.

Different media permit different degrees of richness in interpersonal communication. Face-to-face communication conveys the most information, since in addition to spoken words a conversation also includes facial expressions, gestures and a host of other nonverbal elements. All nonverbal elements are stripped away in email communication, leaving only words. IM is somewhere in between. That is, while there are still no nonverbal elements, the conversation may occur in real time and so permits the interaction to be paced more like a conversation. There is also a rich slang vocabulary in IM. The ability for the user to evaluate personal elements of the reference interaction – such as the librarian's politeness and interest – depends on the richness of the medium, and it may not be possible at all with “thin” media like email.

For desk reference the metric “willingness to return” means willingness to return to ask another question of the same librarian. For online reference, the user often has no control over which specific librarian answers a question. Willingness to return is an appropriate metric of user satisfaction for online reference, but it must be modified to mean willingness to return to submit another question to the same *service*. This metric is an important measure of success. Since there are so many question-answering services online, both based in libraries and not (for example, Yahoo Answers), users may not have the brand loyalty that comes with only having access to one library's reference desk.

The completeness and usefulness of the answer provided by the librarian is an important metric in evaluating any reference service. A common question for a librarian to ask at the end of a reference interaction is something like, “Does this answer your question?” It may not be possible for the user to reliably answer that question, however. Often the user requires some time to make use of the information provided, to apply it in whatever context motivated the question. Information provided in answer to

quick fact and ready reference types of questions may be usable immediately. When a user has a more complex information need, however (related to a research project, for example), the user may not be able to determine if the answer was useful without spending some time applying the information.

It is therefore important to collect data from users on their perceptions of the service at points in time when they can reliably provide that data. Immediately following the interaction users will be able to comment on their initial impressions of the service, which may include such measures as their perception of the librarian's helpfulness and politeness and the ease of use of the software. For a user to provide reliable data on the accuracy, completeness and usefulness of information provided, allowing the user time to read, synthesize and use the information provided may be necessary. Exit surveys at the conclusion of the reference interaction are a common method for collecting data from users, but these should be limited to collecting only data that users can reliably provide immediately following the interaction. Data requiring more time may be collected using follow-up interviews with users, provided some means for contacting the user is also collected.

Conducting an Evaluation of Your Service

Prior to conducting any evaluation, it is crucial to know what individuals or groups are the stakeholders and what criteria will be used to determine success or failure. For evaluations of online services, these two things are often known in advance. Stakeholders may include librarians, library administrators, users, state and local government officials, funding agencies and others. Criteria for success may include volume of questions submitted or answered per unit time, speed with which answers are provided, user satisfaction or other impacts on the user community, the service's reach into new or existing user communities and others.

In the event that the stakeholders or the criteria are unknown, a method that may be used to identify them is evaluability assessment (EA). The purpose of EA is to clarify what is to be evaluated and to whom the evaluation findings will be presented – which often means those individuals and groups in positions to make decisions about the service. These stakeholders may have differing criteria according to which they will consider the service a success.

It is the evaluator's job to balance potentially differing criteria and make recommendations for how the evaluation should be conducted. Done properly, EA can save time and resources, since an evaluation using poorly defined criteria or on a poorly defined service, like any poorly defined research project, is all too likely to answer the wrong questions, to identify findings that are not useful or meaningful to the audience or to run over budget.

Once the criteria according to which a service will be evaluated are clear and the audience for whom the evaluation is being performed is identified, all that remains is to perform the actual evaluation. Of course, that task is easier said than done. However, if an EA was performed, then some methods for collecting and analyzing data are likely to have already been identified. The evaluator must then decide what the best instruments are for collecting the desired data, whether surveys, interviews or web log statistics. As discussed above, there is a wide range of possibilities.

When presenting evaluation results, the evaluator should only present those that specifically address the evaluation criteria and that are most relevant to the stakeholders. For people who enjoy doing the work of research and evaluation, having a lot of data is exciting. It is easy for evaluators to forget that not everyone loves looking at data as much as they do. Audiences can easily be overwhelmed by too much data. It is therefore up to the evaluator to be selective about what to present. Evaluators should not be biased, presenting only those results that support some agenda. Neither should evaluators provide only the results that stakeholders want to hear. Rather, evaluators should provide those results that will enable stakeholders to make decisions about the service, based on the specified criteria for success.

It is also usually worthwhile to present recommendations in an evaluation report. Recommendations may address possible solutions to problems with the service, suggestions for decisions that are pending about the service, future directions for the service or any number of other topics. The evaluation results and the concerns of the various stakeholder groups will dictate what recommendations might be appropriate. Stakeholder groups are not obligated to accept the evaluator's recommendation, of course, and often do not. But it is often useful for stakeholders to have

POMERANTZ, continued

recommendations from the evaluator – as someone in possession of a great deal of information about the service – to inform their decision making.

It is important to remember that research and evaluation are not the same thing: research is what you do to make evaluation possible; evaluation is one possible use for research findings. The purpose of research is to learn about a specific thing; the purpose of evaluation is to come to a judgment about that thing. When that thing is an online reference service, there are many parts that may be researched separately and about which judgments

may be made. And, as with any service, those judgments will be informed by the different concerns of different stakeholders. Libraries exist at present (perhaps always) in an environment of tight budgets, which requires all expenses to be justified. Reference work, and particularly online reference, is potentially an expensive service to offer. It is therefore critical that evaluations of online reference services be conducted well, so that they are useful to stakeholders in making decisions that can affect the very future of the service and the library itself. ■

Resources for Further Reading

- Elliot, D. S., Holt, G. E., Hayden, S. W., & Holt, L. E. (2006). *Measuring your library's value: How to do a cost-benefit analysis for your public library*. Chicago: American Library Association.
- McClure, C. R., Lankes, R. D., Gross, M., & Choltco-Devlin, B. (2002). *Statistics, measures and quality standards for assessing digital reference library services: Guidelines and procedures*. Syracuse, NY: Information Institute of Syracuse. Retrieved October 14, 2007, from <http://quartz.syr.edu/quality/>
- Neuhaus, P. (2003). Privacy and confidentiality in digital reference. *Reference & User Services Quarterly*, 43(1), 26-36.
- Nicholson, S. (2004). A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of Documentation*, 60(2), 164-182.
- Nicholson, S., & Smith, C. A. (2007). Using lessons from health care to protect the privacy of library users: Guidelines for the de-identification of library data based on HIPAA. *Journal of the American Society for Information Science and Technology*, 58(8), 1198-1206.
- Pomerantz, J., & Luo, L. (2006). Motivations and uses: Evaluating virtual reference service from the users' perspective. *Library & Information Science Research*, 28(3), 350-373.
- Pomerantz, J., Mon, L., & McClure, C. R. (in press). Methodological problems and solutions in evaluating remote reference service: A practical guide. *portal: Libraries and the Academy*, 8(1).
- Radford, M. L., & Kern, M. K. (2006). A multiple-case study investigation of the discontinuation of nine chat reference services. *Library & Information Science Research*, 28(4), 521-547.
- Trevisan, M. S. & Huang, Y. M. (2003). Evaluability assessment: A primer. *Practical Assessment, Research & Evaluation*, 8(20). Retrieved October 14, 2007, from <http://PAREonline.net/getvn.asp?v=8&n=20>