Peer review of chat reference transcripts: Approaches and strategies

Jeffrey Pomerantz[a],[*], Lili Luo[a], Charles R. McClure[b]

[a]School of Information and Library Science, University of North Carolina at Chapel Hill, CB 3360, 100 Manning Hall, Chapel Hill, NC 27599-3360, USA
[b]School of Information Studies, Florida State University, 101 Louis Shores Building, Tallahassee, FL 32306-2100, USA

Abstract

NCknows is a collaborative statewide chat-based reference service, launched in North Carolina in February 2004. The authors were contracted by the State Library of North Carolina as program evaluators for the NCknows project. This article reports on one component of the overarching evaluation effort: an analysis of the transcripts of NCknows chat sessions. This analysis was performed as a peer review, where NCknows librarians critiqued transcripts from the service. This study has three main findings: (1) The quality of reference service being provided by NCknows is high overall. (2) NCknows librarians are more engaged with users than are 24/7 staff members, but they are no more skilled in research or use of information sources. (3) Public librarians provide superior service, but academic librarians provide superior referrals. The implications of these findings for staffing chat reference services are discussed, with respect to librarians' credentials and the participating libraries' service philosophies.
© 2005 Elsevier Inc. All rights reserved.

* Corresponding author.
   E-mail address: pomerantz@unc.edu (J. Pomerantz).

1. Introduction

The evaluation of library reference services has probably been conducted for as long as reference services have been offered in libraries. Evaluation of reference services began in earnest in the late 1960s and early 1970s, when tight budgetary situations required all services to justify their place in the library. The same situation applies today, with tight budgets requiring a close examination of all services offered in a library. The need for such evaluation is especially crucial for new services such as chat reference.

Some early work in the evaluation of library reference services included analyses of the cost of providing reference service. A number of early studies arrived at surprisingly high cost-per-transaction figures (Murfin, 1993). These high figures in part led to other evaluation measures than cost being sought by reference services seeking to justify their existence in the library. One widely used alternative measure is the quality of the service being provided. While this measure is popular, it is also difficult to operationalize and implement. Take, for example, Hernon and McClure's (1986) and Saxton and Richardson's (2002) studies of reference quality. Both studies considered a question to be answered accurately that was answered completely and correctly, and not referred to another librarian or agency. Hernon and McClure, on the one hand, developed a set of quick fact and bibliographic questions specifically for their study and had proxies ask these questions at reference desks in public and academic libraries. Their findings indicated that reference librarians answer only 55% of the questions submitted to them correctly. Saxton and Richardson, on the other hand, used a random sample of actual questions asked by users at reference desks rather than artificial questions and found answer accuracy rates of over 90%. As will be discussed below, this study implemented the assessment of answers differently that either Hernon and McClure or Saxton and Richardson, but operationalized answer accuracy similarly.

Saxton and Richardson (2002) present an excellent overview of the many measures that have been used in the evaluation of reference services (pp. 35–44), in addition to answer accuracy. Many of the measures proposed for evaluation of desk reference apply in the online environment as well. Work has been done recently that offers statistics, performance measures, and quality standards to assess virtual reference services (McClure, Lankes, Gross, & Choltco-Devlin, 2002; Virtual Reference Desk Project, 2003), and Hirko (2004) goes so far as to offer checklists and methodologies that may be utilized in such assessments. In short, there is no lack of proposed approaches and measures to assess reference services in general and chat reference in particular. One approach that works especially well for the evaluation of chat reference is the unobtrusive evaluation methodology employed by many evaluations of desk reference services, including the ones discussed above. Unobtrusive evaluation is a "secret shopper" methodology where the researcher or a proxy asks a question as a user, so that the librarian does not know that he or she is being observed. The difference between unobtrusive evaluation of desk and virtual reference services, however, is that most applications that are used for chat reference store the transcripts of the sessions. With access to an archive of chat reference sessions, it thus becomes possible to conduct unobtrusive evaluation without the evaluator posing as a user, that is, with real questions from real users.

2. The NCknows service

NCknows, a chat-based reference service launched in North Carolina in February 2004 (www.ncknows.org), is a collaborative statewide service, which as of this writing has completed an 18-month pilot phase.[1] During this pilot phase, 19 libraries of a range of types (academic and public; in research universities and community colleges, urban and rural, large and small) participated in the project and librarians in these libraries staff the service.[2] Plans are under way to involve more libraries in the service and 20 or more have joined NCknows when the pilot phase was completed in June 2005.

The NCknows service utilizes the 24/7 Reference software tool set to provide chat-based reference service (www.247ref.org).[3] The 24/7 Reference application was developed specifically for use in library reference and allows a set of functionalities common to many commercial applications for chat-based reference: instant messaging (IM), graphical co-browsing, Web page and document pushing, customization of prescripted messages, storage of transcripts, statistical reporting, and other functions.

24/7 Reference is, confusingly, the name both of the software tool set and the company that provides it. One of the services offered by the company 24/7 Reference to library licensees of the application 24/7 Reference (hereafter just "24/7") is membership in a distributed network. This network has two components: other libraries that license 24/7 and employees of 24/7 Reference-the-company. Each library that licenses 24/7 has its own "queue," which uniquely identifies the library's Web site via which a user connected to the service. If a library becomes a member of the 24/7 network, other libraries or 24/7 Reference employees may "pick up" questions that are received via a library's queue, during the hours that the library is not offering service, or if the library receives an "overflow" of questions: that is, more questions coming in than can be handled by the library at the time. Some, but not all of 24/7 Reference's employees staffing this network are librarians, and many are paraprofessional librarians and library science students. When a library licenses 24/7, the library can choose to become a member of the network that includes 24/7 Reference's employees only, or the larger network of other libraries that license 24/7, or the library can choose to "stand alone." The NCknows service, during the pilot phase, was a member of the network that includes 24/7 Reference's employees only. Thus, any users' questions that were not answered by librarians in libraries participating in NCknows were answered by employees of 24/7 Reference.

Chat- and e-mail-based reference services frequently have policies on their Web sites that indicate the topics on which the service will answer questions, and the depth of service that is offered. NCknows' policy concerning the topics and depth of service offered can be found at

---

[1] The NCknows project is funded by a Library Services and Technology Act (LSTA) grant and is being coordinated by the State Library of North Carolina's Library Development Section (http://statelibrary.dcr.state.nc.us). The authors have been contracted by the State Library of North Carolina as the evaluation team for the NCknows pilot project.

[2] The list of participating libraries can be found on NCknows' Web site at http://www.ncknows.org/partlib.htm.

[3] In August 2004, the 24/7 Reference application was acquired by OCLC (www.oclc.org), and the current plan is for 24/7 Reference to be merged with the QuestionPoint service administered by OCLC. For more information on this acquisition, see Quint (2004).

ncknows.org/using.htm. The NCknows service offers reference assistance on a wide range of topics; indeed, the only topics on which NCknows explicitly states that it does not offer assistance is legal and medical questions. NCknows also offers a range of depth of service, from answering simple factual and statistical questions, to assistance with research, to book recommendations.

The primary motivation behind this evaluation effort is to determine whether or not collaborative chat-based reference is an effective way to meet the information needs of North Carolinians. A number of secondary evaluation questions also motivate the evaluation effort, including such questions as how can the quality of the service provided be measured, what is the value of collaboration in a chat-based reference service, what is the value added if different types of institutions work together, and is virtual reference expandable to the whole state?

The evaluation of NCknows is a mixed-method effort, utilizing both quantitative and qualitative methods. To answer the evaluation questions, the evaluation effort analyzes NCknows from the perspectives of the three stakeholder groups involved with and impacted by the service: library users, the individual libraries and librarians participating in NCknows, and the entire collaborative effort. The evaluation effort is employing several different data collection methods to answer the various evaluation questions.

This article reports on one component of the overarching evaluation effort: an analysis of the transcripts of NCknows chat sessions. This analysis allows evaluation from the point of view of the NCknows collaborative as a whole: Critiquing transcripts enables analysis of the quality of reference service being provided to users, under the banner of NCknows. Three evaluation questions were developed that are answered by this transcript review:

1. What is the quality of reference service being provided to users by NCknows?
2. How does the quality of reference service provided by 24/7 staff members compare to the quality of service provided by NCknows librarians?
3. Can a librarian in a public library provide adequate responses to questions from academic library users, and vice versa?

3. Background

3.1. Overview of chat-based reference

The earliest virtual reference services were offered via e-mail, as outgrowths of existing reference desk services in academic and special libraries (Howard & Jankowski, 1986; Kittle, 1985; Roysdon & Elliott, 1988; Weise & Borgendale, 1986). These virtual reference services were developed largely to extend the hours of availability of the reference desk, by allowing asynchronous interactions between the user and the librarian. Many libraries continue to offer asynchronous reference services, although nowadays most offer the service via both e-mail and Web forms (Goetsch, Sowers, & Todd, 1999; Janes, Hill, & Rolfe, 2001). The percentage of questions submitted to asynchronous reference services via Web forms has in recent years outstripped the percentage submitted via e-mail (Carter & Janes, 2000; Janes et al., 2001).

Leveraging users' familiarity with Web forms, and also the growing use of IM technology (Shiu & Lenhart, 2004), many reference services have, since the mid-1990s, begun offering chat-based reference service. Chat reference is defined here as a computer-mediated reference assistance, and following Francoeur (2001), "where the core of the communication between librarian and user is an exchange of text messages sent in real-time" (p. 190). Several commercial applications exist that are designed specifically for chat-based reference, including the eponymous 24/7 Reference, Tutor.com's (formerly LSSI's) Virtual Reference Toolkit (www.vrtoolkit.net), the Library of Congress and OCLC's QuestionPoint (www.questionpoint. org), and Docutek's VRLplus (www.docutek.com), to name only a few of the most widely used. These applications bear a resemblance to commercial help desk applications, such as L. L.Bean's Live Help (www.llbean.com/customerService/contactUs/) and Lands' End Live (www.landsend.com/cgi-bin/live_direct.cgi). Indeed, 24/7 Reference and the Virtual Reference Toolkit are extensions built on top of eGain's business call center software (www.egain.com). As mentioned above, these applications share similar sets of functionalities, including IM, graphical co-browsing, Web page and document pushing, customization of prescribed messages, storage of transcripts, statistical reporting, and more. For excellent comparisons of the features of these and other applications for chat-based reference, see Hirko (2002) and Ronan (2003, pp. 31–34).

More recently, some libraries have been offering chat-based reference service using IM applications such as AOL Instant Messenger (AIM). Implementing commercial applications for chat reference is fairly expensive (at least for most libraries' budgets), so using AIM for this purpose is a low-cost alternative because AIM may be downloaded for free from AOL's Web site (www.aim.com). Another advantage of IM applications for chat-based reference is that many younger computer users are already familiar with the technology; Janes (2002) suggests that use of IM applications may therefore attract younger users to make use of the reference service.

One major drawback of using an IM application for chat-based reference is that it is a "thin" application: it is possible to transmit only text, hyperlinks, files, and emoticons; the more sophisticated functionality offered by commercial applications for chat reference is not available. Two important functions here are the ability for multiple librarians who may be geographically dispersed to monitor the same "queue" of incoming questions from users, and for a librarian to transfer a chat transaction to another librarian. Leveraging these two functions, several consortia of chat reference services have been formed since the late 1990s. Some of these consortia are composed of libraries within a single state or geographic region, such as QandA-NJ, a service of the New Jersey Library Network (www.qandanj.org), KnowItNow, the Cleveland Ohio public library's CLEVNET Consortium (www.knowitnow24x7.net), and the Western New York Library Resources Council (www.wnylrc.org). Some of these consortia are composed of libraries using the same software application, including those discussed above. The NCknows service is both composed of geographically collocated libraries (within North Carolina), and libraries using the same software application (24/7 Reference).

A possible downside of a library joining a consortium of chat reference services is that doing so obligates a library to support users from outside its primary user community. It may be difficult to justify the expenditure of resources to support users from outside of the primary

user community to a library's funding agencies. More important to this evaluation effort, however, is the flipside of joining a consortium: one library's users may be supported by other libraries, and these other libraries may be of a different type, for example, public instead of academic. Public and academic librarians generally have different approaches to reference service: public librarians generally try to provide answers to reference questions (what Tyckoson, 2001, refers to as the "maximum model" of reference service), whereas academic librarians are more likely to try to engage in bibliographic instruction to teach users how to find answers themselves (Tyckoson's "minimum model") (Conaway, 2000; MacDougall & Quinlan, 2001). This issue leads to the second and third evaluation questions motivating this transcript review.

3.2. Overview of reference evaluation

Evaluation efforts are essential for reference services for a number of reasons. Perhaps most importantly, evaluation provides the library administration and the reference service itself with information about how well the service is meeting its intended goals, objectives, and outcomes; the degree to which the service is meeting user needs; and if resources being committed to the service are producing the desired results. It is this that leads to the first evaluation question motivating this transcript review: as discussed above, the primary motivation behind the evaluation of the NCknows pilot project is to determine whether or not collaborative chat reference is an effective way to meet the information needs of North Carolinians, and part of that effectiveness is the quality of the reference service being provided to users by NCknows.

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 progress and success of the service. Such services cannot be improved if there is no knowledge of the existing problems or deficiencies of the service. While evaluation data are necessary to assist decision makers in managing a reference service, they are not sufficient in and of themselves. All evaluation takes place in a political context in which different stakeholder groups (librarians, 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. Despite the development and implementation of a variety of evaluation measures, different stakeholder groups may interpret evaluation data differently. The data that result from evaluation efforts provide baseline information that can inform decision makers as they discuss the activities and results of the service. Ultimately, it will be the evaluation team's task to make recommendations to NCknows about changes and improvements that may be made to the implementation of the NCknows service.

4. Methodology

All chat transcripts from the launch of the NCknows service on 16 February through 30 June 2004 made up the pool from which transcripts were sampled for this study. This pool

includes 3412 transcripts. All chat sessions are ended with a "resolution code," which is intended to indicate the manner in which the chat session was concluded. Fig. 1 presents the 12 resolution codes that exist as of this writing.[4] All transcripts closed with the resolution code "Test" were removed from the pool ($n = 336$). The sampling pool thus included a total of 3076 transcripts. Nearly 10% of all transcripts are therefore closed with the resolution code "Test," although the raw number of transcripts closed with this code decreases month by month. Unfortunately, participating librarians did not use resolution codes consistently and some training sessions were not closed with the Test resolution code. Because there is no way to automatically filter these transcripts out of the sampling pool, it was later discovered that seven test transcripts had been included in the set of transcripts that were sent to librarians to be evaluated. The issue of resolution codes will be discussed further in the discussion section.

The investigators divided the pool of 3076 transcripts first into two pools: those in which the librarian was in one of the libraries participating in NCknows (1378 transcripts), and those in which the librarian was an employee of 24/7 (1698 transcripts). NCknows librarians thus handle approximately 45% of the sessions that come in to the NCknows service, and approximately 55% are handled by the 24/7 employees.

The pool of transcripts handled by NCknows librarians was then subdivided into two groups: those in which the librarian was in an academic library (699 transcripts, or 51% of transcripts handled by NCknows librarians) or a public library (679 transcripts, or 49%). There were thus three pools of transcripts: NCknows academic (699 transcripts), NCknows public (679 transcripts), and 24/7 (1698 transcripts). As stated above, NCknows offers reference assistance on a wide range of topics and depths of service. These three pools were composed of questions and answers on a range of topics and at a range of depths; these pools of transcripts were completely natural, as they occurred "in the field," as it were, of the NCknows service.

Before the launch of NCknows, libraries applied to be involved in the pilot phase of the project. In these applications, each library specified two librarians who would participate in NCknows. As mentioned above, 19 libraries participated in NCknows during the pilot phase of the project. Thus, 38 librarians were involved in NCknows from its inception; that is, from the planning stages, through the launch of the service, and on past the transcript review (most of these librarians are still involved as of this writing). These 38 librarians were the reviewers for this transcript review. Additional librarians in some of these 19 libraries were trained and started to provide chat reference service between the service's launch and the transcript review, but these librarians were not included as reviewers, as it was desired to have the reviewing librarians all have the same duration of experience with NCknows.

Twelve transcripts were randomly sampled per reviewing librarian from these three pools of transcripts, so that all librarians reviewed four transcripts from each pool. No librarians reviewed transcripts in which they or any colleague in the same library was the

---

[4] For further discussion of resolution codes, see 24/7's *Quickstart Manual* section on Ending the Session: www.247ref.org/quickstart/.

> **COMP:** Chat session completed successfully
> **INAPPROPRIAT:** Original question was inappropriate for the service, obscene, a prank, etc.
> **TECHPROB:** The session was not successfully concluded due to connection or other technical problems
> **TEST:** The session is a training session or a test.
> **WCS-ABAN:** The user abandoned the session before a librarian connected. This code is automatically assigned by 24/7 application.
> **WCS-CLOS:** The session was not closed properly by a librarian. This code is automatically assigned by 24/7 application.
> **WCS-GONE:** The librarian attempted to connect but was unable to due to technical problems. This code is automatically assigned by 24/7 application.
> **XFERIN:** The session requires a followup and an e-mail message is referred to another participating library.
> **XFERNCACADEM:** A referral is made to an academic library.
> **XFERNCLEGAL:** A referral is made to a law library.
> **XFERNCPUBLIC:** A referral is made to a public library.
> **XFERNCSPEC:** A referral is made to a special library.

Fig. 1. Resolution codes and their meanings.

librarian. Given 19 libraries, 2 librarians per library, and 12 transcripts per librarian, a total of 456 transcripts were sent to librarians to be reviewed.

The number 12 was chosen arbitrarily as being large enough that a pool of transcripts would result from which statistical significance could be computed, but small enough that conducting the review would not place too great a burden on the librarians' time. (As it turned out, reviewing 12 transcripts took librarians between 20 and 40 minutes.) Twelve was also selected because it is divisible by three, and thus librarians could be made to review transcripts from all three pools of transcripts described above. The transcripts handled by NCknows librarians were overrepresented in this sampling (8 out of the 12 transcripts reviewed by each librarian) because this evaluation effort was intended to be primarily an evaluation of the quality of service provided by NCknows librarians. Transcripts handled by 24/7 employees were also reviewed, however, because it was important also to evaluate the quality of service being provided by 24/7 under the banner of NCknows.

The 456 transcripts sampled for this review were anonymized by replacing every instance of the user's username with "User" and every instance of the librarian's username with "Librarian." This was necessary to preserve the users' privacy, as well as to insure that the librarian reviewing a transcript would not know who the librarian was whose chat session he or she was reviewing.

The review itself took the form of a Web survey, which was implemented via SurveyMonkey (www.surveymonkey.com), a Web-based survey creation tool. The review consisted of 13 multiple-choice questions and one open-ended question, which can be seen in the subsequent section. SurveyMonkey was used for this review because it automatically performs statistical analyses of responses to closed-ended questions, and because it allows logic to be built into surveys. This logic was used for the review, so that depending on the answer to previous

questions, subsequent questions may be skipped. For example, if the reviewer answered the question in the review "The question is answered completely and correctly" with the option "No answer is provided," then the follow-up question "The user is given one or more specific, authoritative source(s) to support the librarian's answer" was skipped.

A packet of 12 anonymized transcripts was mailed to each of the 38 librarians participating in the review, with a set of instructions. These instructions included all of the questions in the review and suggested that the librarians read through the questions before beginning the review, so as to have the evaluation criteria in mind while reading the transcripts. The instructions also suggested that the librarians read a transcript through completely first, and then start the survey and conduct the review of the transcript, so as to have the full transcript in mind during the review. There was, however, no way to insure that the reviewing librarians conducted the review in the manner suggested. Finally, the instructions included a glossary defining the variables and the criteria according to which the reviewers were asked to review the transcripts.

Given their geographic distribution, it was not possible to train the librarians before the review. All of the librarians participating in NCknows, however, had attended a training session on the use of the 24/7 Reference application, which included a component on conducting chat sessions. Consequently, all reviewers came into the review with similar ideas about how to conduct a good chat reference transaction and what constitutes good chat etiquette. Additionally, providing the reviewers with instructions and the full text of the questions in the review was an effort to insure that the reviewing librarians all used the same criteria in the review.

Every librarian reviewed a unique set of 12 transcripts; in other words, each of the 456 transcripts was reviewed by one and only one librarian. For the purposes of this evaluation, coverage of the maximum possible number of transcripts was the most important consideration, more important than having multiple librarians rate the same transcript. This approach was taken because a primary objective of the evaluation effort is to produce findings that may be used to improve or fine-tune the NCknows service, and the greater the number of transcripts that are reviewed, the greater the chances of discovering issues that need addressing by the service.

The review that each librarian performed on their set of 12 transcripts was an unobtrusive evaluation. As discussed above, the manner in which unobtrusive evaluation has been performed at reference desks is that the researcher or a proxy asks a question as a user, so that the librarian does not know that he or she is being observed. Because the 24/7 Reference application stores completed chat session transcripts, it was possible to conduct an unobtrusive evaluation with real questions from real users.

The Hawthorne effect is of possible concern for this study. First identified by Franke and Kaul (1978), the Hawthorne effect is the impact that being observed has on job performance. As might be expected, when one is being observed one's job performance improves. One common explanation for why this occurs is that being observed provides extra motivation that increases job performance. It seems possible that the Hawthorne effect would impact the results of this study as well, that the fact that the NCknows librarians knew that chat sessions would be reviewed motivated them to provide better service than they might otherwise do. The authors do not believe that this is realistic, however, for three reasons. First, the Hawthorne effect applies only to job performance at the time that the subjects are being observed. This is not the situation

with the transcript peer review, which was an unobtrusive evaluation. Librarians are not observed during chat sessions; it is only after the conclusion of a session that the transcript is reviewed. Second, while only a small subset of all NCknows transcripts were selected for the transcript peer review, all transcripts are reviewed by the NCknows systems administrator within a day or two of being completed. The administrator reviews all transcripts looking for evidence of software problems, as well as for examples of difficult patrons or difficult questions. Finally, the chat transcripts sampled for the review were selected randomly from all of those conducted over a span of 5 months, so the librarians would have no way to know which of their sessions would be later reviewed. For these final two reasons, therefore, even if the Hawthorne effect does impact NCknows librarians' performance, it will impact their performance all the time, thus making that increased level of performance the norm and not the exception.

5. Results

Of the 456 transcripts sent out for review, 429 (94.1%) were reviewed. This remarkably high response rate is due to the fact that the State Library of North Carolina mandated that all librarians involved in NCknows must participate in the evaluation efforts. The response rate was not 100% because of three issues:

1. The inclusion of the seven transcripts mentioned above that were not, but should have been, closed with the Test resolution code. These seven transcripts were not reviewed.
2. Two of the reviewing librarians left their jobs during the transcript review and did not complete the review.
3. Some of the reviewing librarians had technical difficulties while attempting to submit reviews to SurveyMonkey.

Not every question in the review was answered for every transcript, as can be seen in the Ns for the tables below. Some questions were not answered due to the logic built into the survey, so that some questions were deliberately skipped. Some questions may also have been skipped because of technical problems or because the librarian thought that the question was not appropriate for the specific transcript he or she was reviewing.

Table 1
Question 1: The question is answered completely and correctly ($N = 428$)[a]

Response	Percentage of responses
Complete and correct	52.3
Incomplete but correct	26.6
Incomplete and incorrect	4.0
Incorrect	0.9
No answer provided	16.1

[a] The percentages in Table 1 sum to 99.9%. The percentages in some other tables sum to 100.1%. This is due to rounding all percentages to one decimal place.

5.1. One-question analyses

The results from all 13 multiple-choice questions in the transcript review are presented here. These multiple-choice questions all had a "Comments" field, the results from which will be addressed below. There was also one open-ended question in the transcript review from which the results are not presented here: What could the librarian have done differently to improve the chat session? The results from this question will be addressed below, with the comments from the other questions. Overall, the evaluations were very positive for all of the multiple-choice questions.

The first question in the transcript review was "The question is answered completely and correctly." As shown in Table 1, while less than 55% of questions were answered completely *and* correctly, nearly 80% of questions were answered correctly, if not completely. For the transcripts in which no answer was provided, the reviewing librarians' comments indicated several reasons why this was so: technical problems prevented the successful conclusion of the session; the question was inappropriate for the service, which included obscene and otherwise inappropriate questions, as well as those that were appropriate for another service and the librarian referred to appropriately; and the librarian in the chat session promised to follow-up with the user by e-mail.

The following four questions in the transcript review were as follows:

1. The user is given one or more specific, authoritative source(s) to support the librarian's answer.
2. The librarian promises to follow-up with the user by e-mail, telephone, or some media other than chat.
3. The user is referred to an appropriate other service.
4. The librarian told the user what s/he was doing as the transaction progressed.

As mentioned above, SurveyMonkey allows logic to be built into surveys. This logic was used for these questions, so that depending on the answer to one question subsequent questions may be skipped. If the reviewing librarian answered the first question in the review as "no answer provided," then Question 2 was skipped, as there were no sources provided that could be authoritative. If the reviewing librarian answered Question 1 with "complete and correct," then Question 3 was skipped, as there was no need for a follow-up. The results of Questions 2–5 in the review are presented in Table 2.

Table 2
Review Questions 2–5

Response	Question 2. Authoritative source ($N = 226$)	Question 3. Follow-up ($N = 203$)	Question 4. Referral ($N = 428$)	Question 5. Communication with the user ($N = 428$)
Yes	89.8	31.0	52.9	70.6
No	10.2	69.0	47.1	29.4

While nearly 80% of questions in chat sessions were answered correctly, if not completely, librarians in chat sessions provided authoritative sources for nearly 90% of questions. For the transcripts in which the reviewing librarian indicated that an authoritative source was not provided, the source may in fact have been authoritative, but there may have been no way for either the user or the reviewing librarian to evaluate the quality of the source because the librarian in the chat session did not provide a citation or link to the source for the information provided to the user.

If a librarian cannot answer the user's question in the chat session or provide resources, then the librarian should at least offer to follow-up with the user after conducting some further research off-line. Where the librarian does not do so, it was either because the question was answered completely and so no follow-up was necessary, or the user did not provide an e-mail address and so it would not have been possible for the librarian to follow-up.

If the librarian did not provide a user with a complete and correct answer in a chat session, or follow-up at a later time with an answer, then the librarian should refer the user to another reference service or information resource of some type. Where the librarian does not do so, frequently the reviewing librarian indicated that it was because the question was answered completely and so no referral was necessary. Interestingly, in their comments, many reviewing librarians mentioned specific information resources rather than other services that the librarian should have used in the chat session.

Another component of the overarching evaluation of NCknows is a series of interviews with the participating librarians. One finding that has emerged from these interviews is that the participating librarians believe that one of the most important objectives for a librarian in a chat session is to create rapport with a user and keep the user "in the loop": to communicate often about what the librarian is doing both online and offline, so that the user does not feel either overwhelmed by the functionality chat application or ignored when the librarian does not provide any resources for a few moments. The results from Question 5 indicate that the librarians participating in NCknows are doing well in terms of communicating with users. For the transcripts in which the librarian did not communicate with the user as the transaction progressed, it was because the librarian sent resources to the user without any explanation of the authority or provenance of those resources or why the librarian selected them.

Whether the question was answered completely and correctly or not, if the librarian provided any information resources at all, they should be appropriate to the user's information need. That means not only topically relevant, but authoritative, unbiased, and useful to the user. Table 3 indicates that most librarians selected information resources well: Over 85% of librarians in the chat session provided resources that the reviewing librarian thought appropriate. For the transcripts in which the librarian provided no appropriate information resources or only some, the reviewing librarians' comments indicated that this was because the librarian in the chat session did not indicate what information resources he or she was using, or perhaps relied on personal knowledge. The information resource or resources that the librarian in the session used may therefore have been perfectly appropriate, but because it was not cited there is no way to evaluate it.

Table 3
Question 6: The librarian used appropriate information resources to answer the question ($N = 428$)

Response	Percentage of responses
Yes: I don't know any resources that the librarian could have used that he or she didn't.	47.0
Mostly: The librarian used many or most of the resources I would have used to answer this question.	26.4
Some: The librarian used some or a few of the resources I would have used to answer this question.	12.9
No: The librarian didn't use any of the resources I would have used to answer this question.	13.8

Creating rapport with a user is what reference librarians do all the time in the reference interview to facilitate the elicitation of information from the user so as to figure out how best to meet the user's information need. To accomplish this, librarians are trained to ask the user open-ended questions. Table 4 indicates that such open-ended questions are usually, but not always, needed in a chat session. The fact that librarians themselves stated that open-ended questions were not needed in certain reference transactions raises the question of whether expressing an information need in writing allows the user to more accurately describe the information need than might be the case at the reference desk, so that a reference interview is unnecessary. Even when open-ended and neutral questions are needed, however, they are not always employed. On the other hand, the librarians' comments indicated that sometimes closed-ended questions were appropriate and sufficient to clarify the user's information need, and sometimes the librarian did nothing to attempt to clarify the user's information need.

The flip-side of the librarian asking open-ended and neutral questions at the outset of and during the reference transaction is for the librarian to ask a closed-ended question at the conclusion of the reference transaction. Librarians are trained to ask a closed-ended question to determine if the question has been answered to the user's satisfaction. Table 5 indicates that the librarians in the chat session frequently neglected this step. For the transcripts in which there is no closed-ended question at the conclusion of the session, the reviewing librarians indicated several reasons. Some of these reasons made it clear that a closed-ended question at the conclusion of the session was probably unnecessary: a closing question was unnecessary due to the session being very short or due to the user

Table 4
Question 7: Open-ended or neutral questions are asked at the outset of the transaction to clarify the information need ($N = 428$)

Response	Percentage of responses
All questions asked at the outset of the transaction are open-ended or neutral.	27.3
Some questions asked at the outset of the transaction are open-ended or neutral.	18.0
All questions asked at the outset of the transaction are closed-ended.	11.0
No questions are asked at the outset of the transaction.	25.7
Open-ended and neutral questions are not needed in the transaction.	18.0

Table 5
Question 8: There is a closed-ended question at the end of the initial interview confirming that the librarian understands the user's inquiry ($N = 428$)

Response	Percentage of responses
Yes	36.7
No	63.3

being very clear about their information need, or the user thanked the librarian before the librarian has a chance to ask a closing question. Some of these reasons, however, indicated the same problem as emerged in the previous question, that sometimes the librarian did nothing and asked no questions to attempt to clarify the user's information need.

Another common component of clarifying the user's information need in a reference transaction is confirming what resources the user has already checked. Table 6 indicates that this clarification is often not needed in a chat session, but even when it is needed it is often not done. It is probably unnecessary to clarify resources already checked when, for example, the user requests a known item: a book, Web site, or other resource to which the librarian has access but the user does not. It is also unnecessary when the patron explicitly states that he or she is just beginning to search for information. As the reviewing librarians indicated, however, it is necessary when the user's question is unclear or when the user indicates that his or her question arose from a specific source, such as a newspaper or television news story.

The librarian should ask a closed-ended question at the conclusion of the reference transaction to allow the user the opportunity to state if the question has been answered to his or her satisfaction. If the librarian does not ask such a question, as was found above, the user may not have the opportunity to give this sort of feedback, positive or negative, to the librarian. This fact is reflected in the "Indeterminate" row of Table 7. A user is, however, not limited to the conclusion of the chat session to provide this sort of feedback to the librarian. For the transcripts in which there is no direct or indirect evidence of the user's satisfaction, however, the reviewing librarians indicated that this was due to any of several causes: the user was impatient, or had unrealistic expectations about the scope of the service offered, or was dissatisfied but too polite to say so. Also, the user may have been satisfied with the interaction with the librarian, but not satisfied with the answer provided, or vice versa.

The user's interaction with the librarian is an important component of the user's satisfaction with a reference transaction. Some aspects of this interaction can only be

Table 6
Question 9: The librarian confirms what other resources the user has already checked ($N = 428$)

Response	Percentage of responses
Yes	16.4
No	48.4
It is not necessary to confirm what other resources the user has already checked	35.3

Table 7
Question 10: There is evidence of the user's satisfaction ($N = 428$)

Response	Percentage of responses
Yes: Strong evidence of satisfaction	27.6
Maybe: Indirect evidence of satisfaction	33.4
No: Strong evidence of dissatisfaction	2.8
Indeterminate: No evidence of satisfaction or dissatisfaction	36.2

addressed by users themselves (such as whether the librarian made the user feel comfortable). Some, however, may be addressed by an outside observer. Table 8 presents the findings from Questions 11 to 13, all of which concern the user–librarian interaction. These three questions were evaluated on a 5-point scale, highest to lowest, as follows:

11. The librarian was courteous: Very courteous, Courteous, Neutral, Rude, Very rude
12. The librarian was patient: Very patient, Patient, Neutral, Impatient, Very impatient
13. The librarian was enthusiastic: Very enthusiastic, Enthusiastic, Neutral, Unenthusiastic, Very unenthusiastic

Certainly courteousness on the librarian's part may go some way towards making the user feel comfortable. For the transcripts in which the librarian was neutral and rude, the reviewing librarians' comments indicated that the librarian was overly brusque, formal, or dismissive, or that the librarian did not make sufficient effort to interact with the user. In the several cases where technical problems with the chat session cut off the patron, the reviewing librarians indicated that the librarian in the chat session should have realized that the problem was technical and not that the patron abandoned the session. For example, a reviewing librarian commented that it would have been better to for the librarian in the chat session to write "are you still connected?" rather than "you aren't answering me?" To be fair however, it is often impossible for either participant in a chat session to tell whether the session was cut short because of technical problems or because the other participant abandoned it. (There were no comments submitted with the single transcript in which the librarian was rated as being very rude.)

Related to the librarian's courteousness is the patience displayed by the librarian. These findings make it clear that librarians displayed a great deal of patience in chat sessions. For the transcripts in which the librarian was neutral, impatient, or very impatient, the comments

Table 8
Review Questions 11–13

Response	Question 11. Courteous	Question 12. Patient	Question 13. Enthusiastic
1: Highest	28.3	26.4	19.9
2	47.4	39.7	32.8
3: Neutral	20.8	25.9	43.1
4	3.3	7.0	3.7
5: Lowest	0.2	0.9	0.5

indicated that the librarian sent links to the user faster than the user could reply to them, or, as discussed above, that the librarian did not conduct any reference interview or ask any questions to clarify the user's information need during the transaction.

The third question in the review concerned with the user-librarian interaction asks about the librarian's enthusiasm. These findings make it clear that librarians displayed less enthusiasm than might be hoped. The comments on the transcripts in which the librarian was neutral or unenthusiastic indicated that librarians in the chat session should not tell the user in the chat session that they are helping other users because doing so conveys a lack of enthusiasm about helping the current user. Unfortunately this recommendation is contrary to the recommendation made above that a librarian should keep the user "in the loop" and should regularly tell the user what he or she is doing both online and offline. There were no comments submitted with the single transcript in which the librarian was rated as being very unenthusiastic.

5.2. NCknows versus 24/7

Every chat transcript is automatically assigned a unique number by the 24/7 application. As mentioned above, the reviewed transcripts were anonymized by replacing the user's and the librarians' usernames with the generic "User" and "Librarian." Similarly, this unique number was not included with the transcripts as they were sent to the reviewing librarians. This was done so that the reviewing librarians would not be able to identify the transcripts that they reviewed. The evaluation team, however, coded each transcript with a different unique number and kept track of which coded number corresponded to which actual transcript number. This was done so that at the conclusion of the transcript review, reviews could be matched to the actual transcripts that were reviewed.

Using this reconstructed data, the transcript review responses were analyzed according to whether a session was handled by an NCknows librarian or a 24/7 staff member. This issue was one of the three evaluation questions that this transcript review was set up to answer. As stated above, some but not all of 24/7's staff members are librarians, and many are paraprofessional librarians and library science students. As a result, there was some debate among NCknows librarians as to whether the responses provided by 24/7 staff members would be as accurate or as complete as responses provided by NCknows librarians. It was therefore important to NCknows to determine whether or not this was in fact the case.

As it turned out, there was little evidence that the responses from NCknows librarians are any better or worse than responses from 24/7 staffers. There was a statistically significant difference between NCknows librarians and 24/7 staffers for only four questions in the transcript review, actually the final four questions in the review. These four questions are presented below. Statistical significance was calculated using the chi-square statistic, with significance computed at $\alpha = 0.05$. The researchers used SPSS software for all calculations in this and the following section, using the raw numbers from the transcript review, although the percentages are presented in Table 9. The chi-square statistic has the drawback that it cannot be meaningfully used when any cell value is low. Because of this, two steps were taken. First,

Table 9
NCknows versus 24/7

	Percentage of responses	
	NCknows librarian	24/7 staff
There is evidence of the user's satisfaction ($N = 428$)		
Yes	32.0	17.6
No	3.0	2.3
Indeterminate	65	80.2
The librarian was courteous ($N = 428$)		
Very courteous	32.3	19.1
Courteous	44.8	53.4
Neutral	19.9	22.9
Rude	2.7	4.6
Very rude	0.3	0.0
The librarian was patient ($N = 428$)		
Very patient	31.6	14.5
Patient	39.1	41.2
Neutral	21.5	35.9
Impatient	6.7	7.6
Very impatient	1.0	0.8
The librarian was enthusiastic ($N = 427$)		
Very enthusiastic	22.3	14.5
Enthusiastic	32.8	32.8
Neutral	40.2	49.6
Unenthusiastic	4.1	3.1
Very unenthusiastic	0.7	0.0

when the values in all cells in a row were equal to zero, that row was removed. Second, when the value of any cell in a row was less than 10, Fisher's exact test was used to compute statistical significance.

Note that the results from all 13 questions in the transcript review are not presented in this section. What is presented here are those four questions for which a significant difference was found between NCknows librarians and 24/7 staff members.

The findings presented in Table 9 indicate that users expressed satisfaction significantly more often in transcripts where they were chatting with an NCknows librarian than in transcripts where they were chatting with a 24/7 staff member ($P = 0.012$). In other words, NCknows librarians elicited spontaneous expressions of satisfaction from patrons significantly more often than 24/7 staff members did.

The findings presented in Table 9 concerning the courtesy, patience, and enthusiasm of the librarian indicate a curious distribution: users evaluated NCknows librarians as very courteous, very patient, and very enthusiastic significantly more often than they used these terms to describe 24/7 staff members ($P = 0.05$, 0.001, and 0.003, respectively), but users evaluated 24/7 staff members as merely courteous and/or neutral significantly more often than NCknows librarians. In other words, users rated NCknows librarians as providing superior service more often than 24/7 staff members, but 24/7 staff members are rated as providing merely service more often than NCknows librarians.

Table 10
Academic versus public libraries

| | Percentage of responses | | | |
| | Via academic library | | Via public library | |
	Academic librarian	Public librarian	Academic librarian	Public librarian
The question is answered completely and correctly ($N = 253$)				
Complete and correct	40.0	62.0	43.1	69.6
Incomplete but correct	36.4	22.5	22.4	23.2
Incomplete and incorrect	3.6	1.4	3.4	2.9
Incorrect	0.0	0.0	5.2	0.0
No answer provided	20.0	14.1	25.9	4.3
The librarian used appropriate information resources to answer the question ($N = 253$)				
Yes	38.2	56.3	43.1	63.8
Mostly	23.6	21.1	22.4	21.7
Some	12.7	7.0	19.0	8.7
No	25.5	15.5	15.5	5.8
The librarian confirms what other resources the user has already checked ($N = 253$)				
Yes	20.0	28.2	6.9	20.3
It is not necessary to confirm what other resources the user has already checked	23.6	42.3	32.8	31.9
No	56.4	29.6	60.3	47.8
There is a closed-ended question at the end of the initial interview confirming that the librarian understands the user's inquiry ($N = 253$)				
Yes	36.4	54.9	27.6	36.2
No	63.6	45.1	72.4	63.8
The librarian was courteous ($N = 253$)				
Courteous	38.2	43.7	41.4	44.9
Very courteous	30.9	46.5	22.4	37.7
Neutral	29.1	7.0	29.3	17.4
Rude	1.8	1.4	6.9	0.0
Very rude	0.0	1.4	0.0	0.0
The user is referred to an appropriate other service ($N = 114$)				
Yes	63.6	25.9	57.6	66.7
No	36.4	74.1	42.4	33.3

5.3. Academic versus public libraries

Using the reconstructed data containing the unique transcript numbers assigned by the 24/7 application, the researchers analyzed transcript review responses according to two criteria: whether a session came into NCknows via an academic or a public library's Web site, and whether a session was handled by a librarian in an academic library or a public library. This issue was also one of the three evaluation questions that this transcript review was designed to answer. As discussed above, public and academic librarians generally have different approaches to reference service. There was some question among NCknows librarians whether "heterogeneous" reference services (where the librarian is a public librarian but the user is an academic library user, or vice versa) can provide adequate responses to users'

questions. It was therefore important to NCknows to determine whether or not such heterogeneous service was in fact adequate. There was a statistically significant difference between public and academic librarians for nearly half (6 out of 13) of the questions in the transcript review. Again, the results from all 13 questions in the transcript review are not presented in this section, but only those six for which a significant difference was found. Statistical significance was again calculated using the chi-square statistic, with significance computed at the 0.05 level.

The findings presented in the first five sections of Table 10 indicates a similar curious distribution to Table 9: public librarians are rated significantly higher on the most positive dimension of these questions, but academic librarians are rated significantly higher on some of the less positive dimensions.

The findings presented in the last section of Table 10 (referral to an appropriate other service) does not support the trend displayed in the rest of that table: instead of public librarians being rated significantly higher on the most positive dimension, academic librarians are rated significantly higher for making appropriate referrals. In other words, the service provided by public librarians was rated as being superior to the service provided by academic librarians, but the referrals made by academic librarians were superior to those made by public librarians.

6. Discussion

Hernon and McClure's (1986) and Saxton and Richardson's (2002) studies of reference quality were discussed above. Both of these studies were attempts to measure the quality of reference service being provided. This study too was an attempt to measure the quality of reference service. The differences between these three studies, however, reinforce the point also made above, that while accuracy is a popular measure, it is difficult to operationalize and implement. Service quality was evaluated in Hernon and McClure's study by the researchers themselves. In Saxton and Richardson's study evaluation was performed by both a panel of judges and by users. In this study, the evaluators were the librarians participating in the service. Thus, in these three studies alone, four different stakeholder groups are represented as the evaluators: a research team, external reviewers recruited by the research team, the librarians who provide the service, and the users who use the service. This study proceeds from the position that the librarians who provide the service are useful as arbiters of the quality of the service. This is not to say, however, that these librarians are the only arbiters of service quality. The difficulty of operationalizing and implementing measures of the quality of reference service may in fact be a benefit, as it insures, indeed requires that a diversity of stakeholders' viewpoints be represented in reference service evaluation.

Some argue that users are the ultimate arbiters of the usefulness of an answer (Dervin, 2003; Durrance, 1989). According to this argument, the quality of the information provided by a reference service may be defined as its usefulness to the user. Another component of the overarching evaluation of NCknows was an investigation of users' satisfaction with the information provided, and their use of this information. One question in this study did,

however, ask about whether there was evidence of the user's satisfaction displayed in the chat session transcript itself. This question was analogous to those studies of e-mail reference sessions that analyze the "thank you" messages that the service receives from users (Carter & Janes, 2000; Mon & Janes, 2003; Weissman, 2004). These studies all have found an approximately 20% "thank you rate." These studies considered "thank yous" to take several forms: not just the words "thank you," but also other expressions of satisfaction with the information provided or the interaction with the librarian, offered spontaneously by the user. The "thank you rate" presented in Table 7 is higher than that from studies of e-mail reference, as Mon and Janes suggest is to be expected in chat reference (p. 3). This supports Durrance's (1989) assertion that "interpersonal variables" are an important component of a library user's satisfaction with a reference transaction (p. 35). One of the attractions of chat technology, as a synchronous medium, as Pomerantz (2005) suggests, is that it enables an interaction between participants in a manner that more closely resembles a face-to-face conversation than an e-mail interaction can do. As a "rich" medium (Daft & Lengel, 1986) – or at least richer than e-mail – chat is richer with interpersonal variables that will affect the user's satisfaction with the reference transaction.

There is a tendency for respondents, when reporting on their satisfaction with a service, to be overly generous, especially when that service is provided by another human being with whom the respondent has had some personal contact. This is a frequent explanation for why reports of user satisfaction with reference services is so high, even when users report that the answer provided by the service was not complete (Durrance, 1989, p. 35; Richardson, 2002). This is probably also the case for this transcript evaluation: the reviewing librarians may have been more generous in their reviews because they knew that they were evaluating their colleagues' transcripts (even though the reviewers did not know which colleagues) than they might have been if they had been evaluating transcripts from an unknown service. This is perfectly natural and to be expected. Indeed, this was anticipated, and in the instructions provided to the librarians was a statement to this effect, and a request to "please be honest" so that accurate data could be collected that could be used to establish a realistic baseline for the quality of the NCknows service. This is, however, easier said than done, and it is possible therefore that the findings of user satisfaction with the service may be artificially high.

A user's satisfaction with the reference interaction is important, but that satisfaction will be cold comfort if the information provided is incomplete or inaccurate. A user may not, however, know if this information is incomplete or inaccurate. Indeed, as Belkin, Oddy, and Brooks (1982) suggest, a user who asks a reference question probably does not know enough to evaluate the answer for completeness or correctness, for if she did know enough to make this evaluation, there would be little reason for her to have asked the question in the first place. As shown above, less than 55% of questions were answered completely and correctly, but nearly 80% of questions were answered correctly, if not completely. This finding splits the difference between Hernon and McClure's (1986) 55% and Saxton and Richardson's (2002) 90% answer accuracy rates. This is particularly noteworthy because this study operationalized answer accuracy similarly to both previous studies but employed a different stakeholder group than either previous study employed to perform this assessment. Another difference between this study and either previous study is the participating librarians. Both Hernon and

McClure (1986) and Saxton and Richardson (2002) conducted their studies in physical libraries, so naturally only librarians employed by those libraries were staffing the reference desk, and could answer questions.

For this study there were effectively two populations answering questions: librarians in the physical libraries participating in the NCknows service, and a group employed by 24/7, to which questions were essentially being outsourced. Recall that users rated NCknows librarians as providing superior service more often than 24/7 staff members, but 24/7 staff members are rated as providing adequate service more often than NCknows librarians. A possible hypothesis to explain this may be, as Bunge and Murfin (1988) proposed and NCknows librarians had suspected, credentials. Bunge and Murfin found that paraprofessional librarians rated lower along a range of measures of reference success, as evaluated by patrons, than did professional librarians. Bunge and Murfin found further that in libraries where paraprofessionals were most successful, they handled simpler questions than those handled by professional librarians. Pomerantz (2003) suggests that the issue of credentials has carried over from the world of physical reference into the world of virtual reference services: should only professionally trained librarians be allowed to provide reference service, or may paraprofessionals as well, and for what types of questions? Whitson (1995) refers to these models as "differentiated" versus "undifferentiated" reference service. The training that professional reference librarians have received, and the experience that they have at the reference desk, may indeed make professionals better qualified to provide reference service than paraprofessionals. It is noteworthy, though, that the measures on which NCknows librarians were rated higher were those measures concerned with the librarian's engagement with the user, rather than those measures concerned with the librarian's skill in research or use of information sources. Courtesy, patience, and enthusiasm are all positive personality characteristics for a reference librarian to have, and several researchers have found that a user's satisfaction with a reference transaction is as or more dependent on the interaction with the librarian than with the actual accuracy or completeness of the answer provided by the librarian (Dewdney & Ross, 1994; Durrance, 1989; Gross & Saxton, 2002). This study's findings fly in the face of Bunge and Murfin's findings that paraprofessionals lacked the skills to answer complex reference questions: the results of this study indicate that paraprofessionals possess all of the requisite skills to answer questions and use appropriate information sources, but professionally trained librarians are superior in their display of desirable personality characteristics during the reference interaction.

Similar to the split between NCknows librarians and 24/7 staff, there was a split in this study's findings for public and academic librarians. Recall that the service provided by public librarians was superior to the service provided by academic librarians, but the referrals made by academic librarians were superior to those made by public librarians. A possible hypothesis to explain this may be the service philosophy in public and academic libraries. Another component of the overarching evaluation effort is interviews with the librarians participating in NCknows. In these interviews, several librarians suggested that there should be policies regarding whether NCknows should provide users with answers or resources. This is determined largely by the question itself (whether it is answerable at all), the policies at a particular library, and by the librarian's personal style. Part of an academic library's mission is

to support education in their institution, and part of this means encouraging students to do their own research. Thus, it is likely that an academic librarian may only partially answer a question and provide some but not all resources to answer a question in an effort to get the user to conduct some of their own research. It seems less likely that a public librarian would have this attitude (e.g., Saxton & Richardson, 2002; White, 1997).

In light of this difference in service policy, it is interesting that public librarians were rated higher than academic librarians, and by both public and academic librarians. Given academic libraries' service policies or providing resources or instructions rather than answers, it might be expected that academic librarians would rate chat sessions in which the librarian provided resources or instructions more highly than sessions in which the librarian provided answers, but this was not the case. It would appear that librarians, regardless of their library's service policy, consider providing answers to be the reference service ideal. This said, however, in their comments, the reviewing librarians made it clear that it was also important that the librarians in the chat session provided the user with a variety of information resources *in addition to* an answer, as well as taking the time to provide bibliographic instruction in the use of whatever resources they provided. Thus, it seems that the ideal service philosophy for chat reference is a combination of that from both public and academic libraries: to not only provide answers but also information resources.

Every closed-ended question in the transcript review was followed by an open-ended Comments field. Based on these comments, there are several recommendations that may be made for librarians offering chat reference service.

First, librarians should pay attention to which library's Web site a user used to access NCknows. This allows the librarian to determine what resources the user will have access to locally because different libraries will have different print resources as well as different subscriptions to online resources. Knowing a user's "home" library also allows the librarian on the other end of a chat session to forward a follow-up question for the user to that library, if necessary.

Second, and perhaps most importantly, knowing how a user came into the service allows a librarian to know which library the user probably thinks he or she is contacting. This is important because users, particularly first-time users, may not be aware that the librarian with whom the user is chatting may not be in their local library, or even in their state. The Web site of a chat reference service should clearly explain this fact about the service.

Third, the reviewing librarians made it clear in their comments that explanations for the user are one of the most important aspects of the chat reference transaction, both before the session begins and during the session itself. The page-pushing and co-browsing features of chat reference applications are unfamiliar technologies for many users, so it is especially important for librarians to explain what they are doing when using these features. Users also have no way of knowing when the librarian is searching for information offline without pushing pages to the user, so it is important for the librarian to take the time to explain the procedures that he or she uses.

These recommendations are actually little different than the behavior that librarians are trained to display at the reference desk. Indeed, the reviewing librarians made it clear

from their comments that behaviors that attempt to reproduce behavior from the reference desk were rated higher. The reviewing librarians made it clear that a deliberate effort by the librarian to compensate for the comparative lack of "richness" (Daft & Lengel, 1986) in chat interactions (compared to face-to-face interactions, e.g., a lack of body language and facial expressions) comprises a good chat session. Such a chat session is one in which there is good communication between the librarian and the user, and the librarian attempts to create the same rapport with the user that can be achieved in person. In other words, the librarian in the chat session should conduct a reference interview: try to clarify the user's question at the beginning of the chat session, ask open-ended and neutral questions to try to elicit more information about the user's information need, ask what resources the user has previously checked, and conclude the session with a question to determine if the question was answered to the user's satisfaction. It is also important for the librarian to be careful with his or her language, as brief statements, lacking non-verbal cues, can be mistaken for rudeness. Finally, a librarian should not be afraid to conclude a chat session with an incomplete answer, but when doing so should be sure to ask for an e-mail address or a phone number to follow-up with the user.

7. Conclusion

The investigators designed three evaluation questions that this transcript review answered. The answer to the first question is that the quality of reference service being provided by NCknows is high overall, as evaluated on the rating scales provided for the questions on the transcript review. The answer to the second evaluation question is that NCknows librarians are more engaged with users than are 24/7 staff members but are no more skilled in research or use of information sources. The answer to the third evaluation question is that public librarians provide superior service, but academic librarians provide superior referrals. A possible explanation for this last finding may be found in the service philosophies of public versus academic libraries.

As with any new service, improvements can be made to NCknows. The reviewing librarians explicitly offered ideas for improvements in their comments, and some improvements can be inferred from similar comments made by several librarians. Some of these recommendations may be implemented as changes to or official policies of the NCknows service, and some recommendations may be implemented informally as guides for librarians' behavior in chat sessions. Policies for quality control will become increasingly important as the NCknows service matures and expands over time.

This transcript analysis allows evaluation from the points of view of the users and the entire NCknows effort: critiquing transcripts enables analysis of the quality of reference service being provided to users, under the banner of NCknows. Reviewing chat transcripts allows for an unobtrusive evaluation of the reference service, with real questions from real users. This enables the findings of the review to be realistic, and the recommendations that may be made to the service to accurately reflect the state of the service.

The evaluation approaches reported here can provide a useful model for others wishing to conduct transcript-based evaluation of chat services. The approach reported here, however, is but one of a number of approaches that the investigators employed to produce a more holistic assessment of the chat-based services of NCknows. The techniques employed in this particular evaluation effort provided useful data that is now being utilized to continue the development of the NCknows service.

Acknowledgments

References

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982, June). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation*, *38*(2), 61–71.

Bunge, C. A., & Murfin, M. E. (1988, March). Paraprofessionals at the reference desk. *Journal of Academic Librarianship*, *14*(1), 10–14.

Carter, D. S., & Janes, J. (2000, Fall). Unobtrusive data analysis of digital reference questions and service at the Internet public library: An exploratory study. *Library Trends*, *49*(2), 251–265.

Conaway, P. (2000, July). One reference service for everyone? *Library Journal*, *125*(12), 42–44.

Daft, R. L., & Lengel, R. H. (1986, May). Organizational information requirements, media richness and structural design. *Management Science*, *32*(5), 554–571.

Dervin, B. (2003). From the mind's eye of the user: The sense-making qualitative-quantitative methodology. In B. Dervin, L. Foreman-Wernet, & E. Lauterbach (Eds.), *Sense-making methodology reader: Selected writings of Brenda Dervin* (pp. 269–292). Cresskill, NJ: Hampton Press, Inc.

Dewdney, P., & Ross, C. S. (1994, Winter). Flying a light aircraft: Reference service evaluation from a user's viewpoint. *RQ*, *34*, 217–229.

Durrance, J. C. (1989, April). Reference success: Does the 55 percent rule tell the whole story? *Library Journal*, *114*(7), 31–36.

Francoeur, S. (2001). An analytical survey of chat reference services. *Reference Services Review*, *29*(3), 189–203.

Franke, R. H., & Kaul, J. D. (1978, October). The Hawthorne experiments: First statistical interpretation. *American Sociological Review*, *43*, 623–643.

Goetsch, L., Sowers, L., & Todd, C. (1999). *Electronic reference service (SPEC Kit No. 251)*. Washington, DC: Association of Research Libraries Office of Leadership and Management Services.

Gross, M., & Saxton, M. L. (2002). Integrating the imposed query into the evaluation of reference service: A dichotomous analysis of user ratings. *Library & Information Science Research*, *24*, 251–263.

Hernon, P., & McClure, C. R. (1986, April). Unobtrusive reference testing: The 55 percent rule. *Library Journal*, *111*(7), 37–41.

Hirko, B. (2002, Fall). Live, digital reference marketplace. *Library Journal*, *127*(17), 16–19.

Hirko, B. (2004). *VET: The Virtual Evaluation Toolkit*. Washington State Library. Retrieved October 27, 2004, from http://vrstrain.spl.org/textdocs/VETmanual.pdf.

Howard, E. H., & Jankowski, T. A. (1986, January). Reference services via electronic mail. *Bulletin of the Medical Library Association*, *74*(1), 41–44.

Janes, J. (2002, October). Live reference: Too much, too fast? *Library Journal*, *127*(17), 12–14.

Janes, J., Hill, C., & Rolfe, A. (2001, November). Ask-an-expert services analysis. *Journal of the American Society for Information Science and Technology*, *52*(13), 1106–1121.

Kittle, P. W. (1985, May). Putting the medical library online: Electronic bulletin boards . . . and beyond. *Online*, *9*(3), 25–30.

MacDougall, H. D., & Quinlan, N. J. (2001). Staffing challenges for a joint-use library: The Nova Southeastern University and Broward County experience. *Resource Sharing & Information Networks*, *15*(1/2), 131–150.

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.

Mon, L., & Janes, J. W. (2003). *The thank you study: User satisfaction with digital reference service*. Retrieved October 27, 2004, from http://www.oclc.org/research/grants/reports/janes/jj2004.pdf.

Murfin, M. E. (1993). Cost analysis of library reference services. *Advances in Library Administration and Organization*, *11*, 1–36.

Pomerantz, J. (2003). Integrating digital reference service into the digital library environment. In R. D. Lankes, S. Nicholson, & A. Goodrum (Eds.), *The digital reference research agenda* (pp. 23–47). Chicago, IL: Association of College and Research Libraries.

Pomerantz, J. (2005). A conceptual framework and open research questions for chat-based reference service. *Journal of the American Society for Information Science and Technology*, *56*(12), 1288–1302.

Quint, B. (2004). *OCLC acquires 24/7 Reference services*. Retrieved October 27, 2004, from http://www.infotoday.com/newsbreaks/nb040823-2.shtml.

Richardson Jr., J. V. (2002, April). Reference is better than we thought. *Library Journal*, *127*(7), 41–42.

Ronan, J. S. (2003). Chat reference: A guide to live virtual reference services. Englewood, CO: Libraries Unlimited.

Roysdon, C. M., & Elliott, L. L. (1988, Fall). Electronic integration of library services through a campuswide network. *RQ*, *28*(1), 82–93.

Saxton, M. L., & Richardson Jr., J. V. (2002). *Understanding reference transactions: Transforming an Art into a Science*. Amsterdam, NY: Academic Press.

Shiu, E., & Lenhart, A. (2004). *How Americans use instant messaging*. The Pew Internet and American Life Project. Retrieved October 27, 2004, from http://www.pewinternet.org/PPF/r/133/report_display.asp.

Tyckoson, D. A. (2001, Fall). What is the best model of reference service? *Library Trends*, *50*(2), 183–196.

Virtual Reference Desk Project. (2003). *Facets of quality for digital reference services*, Version 5. Retrieved October 27, 2004, from http://www.vrd.org/facets-06-03.shtml.

Weise, F. O., & Borgendale, M. (1986, October). EARS: Electronic access to reference service. *Bulletin of the Medical Library Association*, *74*(4), 300–304.

Weissman, S. (2004). *Electronic reference at Morris County Library*. Retrieved October 27, 2004, from http://www.gti.net/mocolib1/mclweb/eref.html.

White, M. D. (1997). Measuring service quality in libraries. In E. Garten, & D. Williams (Eds.), *Advances in library administration and organization* (pp. 1–35). Greenwich, CT: JAI Press.

Whitson, W. L. (1995, March). Differentiated service: A new reference model. *Journal of Academic Librarianship*, *21*(2), 103–111.