Evaluating Reference Service in a Large Academic Library

Cheryl Elzy, Alan Nourie, F. W. Lancaster, and Kurt M. Joseph

An unobtrusive study of the ability of professional librarians to deal with factual questions was conducted at the Milner Library, Illinois State University. Students were recruited to pose questions for which answers were known, to 19 librarians in five departments. In all, 190 test "incidents" (10 questions for each of the 19 librarians) were used. Librarians were evaluated on the accuracy of the responses given and on their responsiveness and helpfulness, as judged by the student proxies. The methods used in the study are described, including the accuracy and attitude scales developed, the major results are presented, and suggestions are made on the follow-up action that seems appropriate after a study of this kind has been performed.

everal investigations have suggested that reference librarians may provide complete and correct answers to factual questions only about half the time.[1] Concern over this disappointing performance and questions about its applicability to reference service at Illinois State University prompted this study.

Illinois State University (ISU), a multipurpose university of more than 22,000 students, offers 191 degree programs in 33 academic departments organized into five colleges. Master's degrees are offered in most areas and doctorates in nine. Milner Library, the central library facility, is organized into five subject divisions on six floors: Education/Psychology, General Reference and Information, Social Sciences/Business, Science/Government Publications, and Humanities/Special Collections. The five divisions are staffed by 20 members of the library faculty, 19 classified employees, and student assistants.

OBJECTIVES

The objectives of the study were (1) to estimate the probability that a user, walking into the library with a factual question, would receive or be led to a complete and correct answer, (2) to determine to what extent student users of the library judge staff members to be responsive and helpful, (3) to identify conditions under which members of the reference staff perform well and conditions under which they perform poorly, and thus (4) to identify ways in which the service might be improved. In other words, the focus of attention would be on the accuracy with which factual ques-

Cheryl Elzy is Head of the Education and Psychology Division, and Alan Nourie is Associate University Librarian for Public Services and Collection Development at Milner Library, Illinois State University. F. W. Lancaster is Professor at the Graduate School of Library and Information Science, University of Illinois–Urbana-Champaign, Urbana, Illinois 61801. Kurt M. Joseph is a Graduate Assistant in the Department of Psychology at Illinois State University, Normal, Illinois 61761. The study described was supported in part by the Council on Library Resources through its Cooperative Research Grant Program.

tions are answered by or with the aid of library staff. While we would agree with such writers as Jo Bell Whitlatch that success in answering this type of question is not the only criterion by which an academic reference service should be judged, we have little sympathy with others who state or imply that accuracy is of little concern to library users, who are more concerned with such things as convenience, timeliness, and the librarian's attitude.[2,3]

The study was not performed as an academic exercise. No hypotheses were formulated. Our main concerns were to get a better idea of the quality of reference service at ISU and, in particular, to identify possible problem areas.

METHODS

The following decisions were made at the outset:

1. The evaluation should be performed unobtrusively. Questions, of the type that might reasonably be put to the various departments of the Milner Library, would be collected and students recruited to pose them to members of the library faculty as though these questions represented their actual information needs.
2. Faculty members would be evaluated on attitudinal characteristics and on whether or not they were able to supply complete and correct answers.
3. The study would be conducted between April 17 and April 24, 1989 to allow sufficient time to recruit and hire students. It would occur during a peak time of activity in the academic calendar—midway between spring break and the end of the semester. Group and individual training sessions for student proxies would be held one week prior to the study. There would be one week of class and one week of finals during which proxies could pose their test questions if they somehow failed to do so during the test period.

Unobtrusive studies of reference service have been performed several times in the last 20 years, and they have been reviewed and evaluated elsewhere.[4] The present study differs from most earlier ones in several ways:

1. In most unobtrusive studies the questions have been posed by telephone rather than by personal visit to the library.
2. Most studies have been performed in order to compare libraries, and perhaps to identify broad categories of factors that might influence reference performance, rather than to derive detailed data on a single library.
3. This study, involving 190 reference transactions, appears to be the largest unobtrusive study of reference service yet to be attempted within a single library.

The director of the library requested that a memorandum be sent to all public service librarians stating that the study would be performed some time during the coming year and that the results would not affect their annual performance evaluations. This was done in January 1989. A few individuals expressed reservations, but after they learned more about the study and its background and intent, their concerns were overcome.

An evaluation form was created for students to complete after each test question had been posed. It recorded the test questions and the answers given by librarians; it also provided space for observations on the attitude and demeanor of the librarians. Each item was followed by a request for open comments from the students.

It was decided that at least 10 questions should be posed to each librarian. More would be desirable, but probably would be too unwieldy. Fewer might not provide a true picture of that librarian's attitude and skills. Therefore, 190 incidents would be recorded by the students—10 questions for each of 19 librarians.

Recruiting students for an unobtrusive study is difficult because researchers cannot simply advertise in the student newspaper or on the bulletin board and still keep the project confiden-

tial. Some contacts were made through professors known to the researchers at ISU and also at Illinois Wesleyan University across town. They were provided a job announcement describing the project and the responsibilities required of students. Some students did apply from both institutions through these personal contacts. In the end, however, the best source for recruitment proved to be Milner Library's own records of students who had applied for positions, but could not be hired for administrative or financial reasons. The person at Milner who is in charge of student employees provided files on 18 ISU students who might be suitable candidates, and these were contacted. Together with two students from Wesleyan, they formed the group of proxies. They were to be paid $75 each for approximately 10 to 15 hours of work, including attending a training session, being interviewed individually by the investigators, asking the questions, filling out the evaluation forms and attending a follow-up, or debriefing, session.

All 20 students were undergraduates: 7 males and 13 females; 4 freshmen, 4 sophomores, 8 juniors, and 4 seniors. Ages ranged from 17 to 23. Although the students came from a wide variety of disciplines, no attempt was made to match reference questions to student majors. The most difficult and time-consuming task was creating the questions to be used. The researchers put together an initial pool of hundreds of questions gathered from reference texts, other reference studies, and the investigators' own backgrounds and experiences.[5] Most were rejected on the first reading. Approximately 150 different questions were selected for research in the Milner Library collection and evaluated for possible use. All questions that could not be answered from Milner's collection were eliminated. Those of a type or subject not usually asked in this library were also discarded. From those that remained, questions were matched by subject and specialty with appropriate floors and librarians, and 58 questions were finally selected. Most were asked more than once. For example, many questions were

asked of each librarian on a particular floor because librarians are responsible not only for reference work in their own subject areas, but for all disciplines housed on their floor (e.g., the music librarian has to answer questions in music, fine arts, literature, and languages). The same question was asked on different floors where it was appropriate to do so. Through repetition of questions, 190 test incidents were created.

The set of questions used differed from that of most earlier studies in that it contained a blend of factual questions (e.g., Who was the secretary of state when Sumner Wells was his assistant?) and of questions that were more research oriented (e.g., I need some articles discussing the short story entitled "The Lottery"). Because all questions could be answered from the resources of the Milner Library, the study was really an evaluation of the librarians' ability to exploit the library's resources.

Simply recognizing that one may be perceived in a certain way by a patron . . . might be enough to solve the problem.

The scheduling of when a student was to pose a question was essential to the project because the student was to seek a particular librarian, identified by desk nameplate. This proved difficult with so many people and questions involved. The researchers wanted to avoid questions being asked twice of the same librarian. Another goal was to spread the questions out evenly during mornings, afternoons, and evenings, and through the several days of the study. Librarians work the reference desks only at certain times, and students had to try to match their schedules with those of the librarians. The researchers also wanted to spread the student proxies out among the five floors and over the evaluation period so that they would not become too familiar to anyone on a given floor and in any time frame. The questions were laid out against the floors with the

librarians' schedules, and 20 question sets of 8 to 10 questions each were created.

The week before the actual study, group training sessions were organized for the student proxies. The background for the project was discussed, as was how the study would be conducted and how the results would be used. Packets were given to each student with a schedule of when and to whom each question should be asked, along with a list of questions and an evaluation form for each. The forms were explained in great detail. Students were encouraged to complete the forms immediately after the encounter so that information and impressions would be fresh in their minds, and comments were strongly encouraged. Finally, their pay, the time frame, and other requirements were discussed. An individual training session was set up with each proxy over the next three days to assist each one in understanding the project, questions, and forms, and to help each construct a "cover story" for the questions in the event of an extended reference interview.

A number of unexpected problems occurred during the study itself, such as librarians being on vacation or out sick, no nameplates on desks, making identification difficult, and so on. But only about 14 questions had to be carried into a third week; all 190 forms were returned, and the students performed diligently. A follow-up session was held with the students a few days after the evaluation period. They were eager to share their thoughts and experiences. Comments ranged from how the time of day affected how librarians handled questions to the perception that the better the students dressed, the better the service received.

The work of verifying and grading answers, and of summarizing the mounds of data collected, continued through summer and fall. The investigators from Milner rechecked each answer on each floor and scored it. The attitudinal scores from each form were compiled, and comments on each librarian recorded. Each individual student, librarian, floor, and question was given a unique code,

and the SPSS program was used to manipulate the data. Demographics on each student were entered, as well as information about each of the 190 incidents (e.g., time of day and date the question was asked, minutes spent with the librarian, and so on).

RESULTS

Each student was asked to record the details of the answer and search results provided by the librarian for each question posed and also to supply information on the source of the answer—title, edition, page number, and so on. Also requested were the date and time of day the question was posed and how many minutes were spent in contact with the librarian. The researchers checked each answer in the 190 cases for completeness and correctness and noted sources used by the librarians. Each incident was assigned a code reflecting the relative success of the librarian in answering the question (see table 1).

Arriving at a viable scoring procedure was difficult. It is relatively easy to score the results of a question posed to a library by telephone. The library can be considered as simply a "black box" and the librarian scored on a binary scale—giving the correct answer or not. For a walk-in to an academic library, scoring is more complicated. It was decided to score from the viewpoint of the student user. Because we were evaluating the role of the library in answering questions rather than in bibliographic instruction, we decided that the best possible result was one in which the user was *given* the complete and correct answer. Anything less should receive a lower score. Being led to appropriate sources by the librarian was judged less satisfactory than receiving a correct answer, but being led to sources was judged better than being pointed to them. It was also thought that being led or pointed to several sources, one of which included the complete and correct answer, was less satisfactory than being led or pointed to only one source that contained the complete and correct answer. Finally, the worst result was one in which the user finished with an incorrect

TABLE 1
SCORING METHOD USED

	Points
Student *provided* with complete and correct answer	15
Student *led to a single source*, which provided complete and correct answer	14
Student *led to several sources*, at least one of which provided complete and correct answer	13
Student *directed to a single source*, which provided complete and correct answer	12
Student *direced to several sources*, at least one of which provided complete and correct answer	11
Student *given an appropriate referral to a specific person or source*, which provided complete and correct answer	10
Student *provided* with partial answer	9
Student given an *appropriate referral to the card catalog or another floor*	8
Librarian *did not find an answer* or suggest an alternative source	5
Student given an *inappropriate referral to catalog, floor, or source*, or librarian unlikely to provide complete and correct answer	3
Student given *inappropriate sources*	2
Student given *incorrect answer*	0

answer. These principles are reflected in the scoring method used.

Some findings from earlier investigations support the scoring method used in this study. Wyma J. Hood and Monte J. Gittings found that librarians accompanied users about 54% of the time when answering reference questions in one academic library.[6] About 46% of the time, they showed them how to find the answer, and about 34% of the time, they found the answer for the user. Thomas Childers, in an unobtrusive study of reference service in public libraries, found that when proxies were directed to library tools, such as indexes or catalogs, they were not accompanied by the librarian in more than half the cases.[7] When directed to browse through shelves, books, chapters, or articles to find answers to their questions, they were not accompanied by the librarian in about 80% (129 out of 159) of the cases. Charles A. Bunge found that library users received correct answers only 47% of the time when merely directed to appropriate sources by a busy librarian. This improved to 59% when a busy librarian helped the users search and to 65% when a librarian not otherwise busy helped the patrons search.[8]

The scoring method we adopted served our purposes, but it can hardly be considered definitive, and other researchers might well disagree with it. Its major limitation is that, to avoid complications in data processing, a zero was assigned to the situation in which a student received an incorrect answer when, in fact, it would make much more sense to give this situation a minus score. If a zero is associated with an incorrect answer, it is necessary to give certain other unsatisfactory outcomes (e.g., an inappropriate referral) a positive value, because they are considered somewhat less heinous, but this hardly seems logical. In fact, a more logical scoring procedure would probably assign a zero to the situation in which no answer was provided (which scored 5 in table 1) and would give minus values to inappropriate referrals and to incorrect answers.

The attitude scale, in contrast to the accuracy scale, was a very simple one. The students rated the librarians on their helpfulness and approachability on a continuous scale of 0 to 10 in answer to

24 questions (e.g., Looks approachable? Acknowledges user's approach to desk? Friendly attitude?). The attitude value for a librarian or a floor is merely the mean of all values recorded by the students for that librarian or floor for all 24 questions.

The scores for the 190 reference incidents are summarized in table 2. In 58 cases, the librarian received a score of 15, meaning that he or she provided a complete and correct answer. However, the philosophy of service in academic libraries often is to provide the appropriate sources or point the student in the right direction. If that level of service is accepted as adequate, then scores of 10 or above would be considered acceptable. This was true of 111 cases (58% of all incidents). If an appropriate referral to the card catalog or another floor is considered acceptable, the library's score increases to 121 cases (about 64% of all incidents). It goes to 128 cases (about 67% of the total) if partial answers are considered acceptable. In 18 of 190 cases (9.5%), the librarian could not or did not find an answer. In 36 (19%) instances,

inappropriate referrals and sources or outright wrong answers were given.

The time of day the question was asked was coded to determine whether it seemed to affect the accuracy or the behavior of librarians. Accuracy decreased during evening hours, but the decrement was not statistically significant. Time of day also had no significant effect on how well the students perceived they had been treated (attitudinal scores).

Table 3 shows how often each of the 58 questions was posed, the accuracy of scores achieved (mean score for questions posed more than once), and the mean time spent per question. Because of minor mix-ups, questions 7 and 17 were not asked. In a few other cases, students failed to record needed data—particularly minutes spent with the librarian—or sufficient information on answers to allow the investigators to verify them. The number of these cases is noted in parentheses in table 3.

The frequency with which a question was asked ranged from 1 to 8 times. Accuracy ranged from 2.3333 on question 43, to 8 total questions scoring 15, and 7

TABLE 2
ACCURACY OF ANSWERS PROVIDED

Answer Code	Frequency	Percent
15	58	30.5
14	24	12.6
13	13	6.8
12	5	2.6
11	8	4.2
10	3	1.6
9	7	3.7
8	10	5.3
5	18	9.5
3	10	5.3
2	16	8.4
0	10	5.3
Missing*	8	4.2
	190	100.0

* Some students failed to provide enough information upon which to base judgments, or asked the question in such a way as to change the expected response, thus invalidating the question.

TABLE 3
QUESTION BY QUESTION RESULTS

Question	Times Posed	Accuracy	Mean # of Minutes Spent on Question
1	2	12.0000	13.5
2	2	13.0000	5.0
3	2	7.5000	4.0
4	2(1)*	15.0000	3.25
5	2	14.0000	6.5
6	4(1)†	5.5000	6.0
7	–		
8	2	8.0000	9.0
9	5	10.2000	4.2
10	2	14.0000	3.0
11	4	9.7500	4.2
12	4	13.2500	8.0
13	2(1)*	14.0000	5.0
14	2	15.0000	3.5
15	2	11.5000	3.0
16	5(1)†	5.2500	4.3
17	–		
18	3	10.5000	15.0
19	5(1)†	10.4000	2.5
20	7	9.7143	8.63
21	4(1)†	13.0000	4.66
22	5	5.8000	6.4
23	7	10.1667	4.93
24	2	8.5000	2.0
25	1	11.0000	2.0
26	6	15.0000	6.66
27	3	11.3333	5.83
28	4	3.4000	4.0
29	4	7.2500	2.5
30	4	7.6667	9.0
31	1	15.0000	12.5
32	3	9.3333	5.6
33	2(2)*		6.0
34	3	15.0000	11.66
35	2	15.0000	7.5
36	3	15.0000	9.3
37	3(1)*	5.0000	6.3

Question	Times Posed	Accuracy	Mean # of Minutes Spent on Question
38	5	11.6000	6.9
39	2	5.0000	12.5
40	3	4.3333	5.3
41	4	7.5000	5.0
42	3	10.3333	7.3
43	3	2.3333	7.83
44	3	7.0000	5.66
45	3	14.3333	3.0
46	2	8.0000	5.25
47	3	6.3333	9.33
48	7	11.5714	9.86
49	8(1)*	7.1429	3.69
50	3	15.0000	4.83
51	6	12.3333	5.25
52	3	10.0000	3.1
53	3(1)*	14.0000	4.3
54	3	6.3333	6.83
55	7	9.5714	6.43
56	2	14.0000	7.5
57	3	12.0000	3.0
58	1	14.0000	5.0

* Missing data in accuracy code.

† Missing data in minutes spent.

scoring 14 on the accuracy scale. Minutes spent went from a low of 0 (the librarian did not leave his or her office, but conducted the transaction through the student assistant covering the desk) to a high of 25 minutes. Mean actual minutes spent went from 2 on questions 24 and 25, to 15 on question 18. There was no appreciable correlation between minutes spent with students and accuracy ($r = .1008, n = 179, p > .09$). Thus, the accuracy variable was collapsed into three categories for further analysis (15 to 10 as acceptable, 9 to 6 as minimal, 5 to 0 as unacceptable), but there was still no relationship between minutes spent and accuracy ($r = .0299, n = 179, p = .346$). Time spent with the student did not reflect how complete and correct the answer was.

Table 4 shows the accuracy and attitude scores achieved by floors or divisions. Floor E achieved both the lowest accuracy figure and the lowest attitude scores. Floor C received highest marks for attitude, while floor B was most accurate.

Table 5 records the attitude and accuracy scores by librarian, along with the number of questions each librarian was asked and the minutes the librarian spent with the students. Attitude scores ranged from 8.75 on a 10-point scale, down to 5.74. Accuracy ranged from 13.889 out of 15, down to 7.125. While the librarian who scored highest in attitude also received highest marks in accuracy, the same was not true of the lowest scores in each category. Both accuracy and attitudinal scores are discriminat-

TABLE 4
ACCURACY AND ATTITUDE SCORES BY FLOOR

Floor	Questions	Accuracy	Attitude
A	30(3)*	10.4074	8.2100
B	30	12.7333	8.2067
C	20(2)*	11.7778	8.5200
D	71(2)*	9.6377	7.7141
E	39(1)*	8.1053	7.1256
Mean	190(8)*	10.1538	7.8342

* Missing data for accuracy scores.

TABLE 5
ACCURACY AND ATTITUDE SCORES FOR EACH LIBRARIAN

Librarian	Number of Questions Asked	Attitude	Accuracy	Mean Minutes Spent
1	10(1)*	8.1900	10.3333	4.35
2	10	7.0000	7.6000	5.45
3	10	7.6300	7.5000	6.975
4	9(1)*	7.6000	7.1250	5.65
5	10(1)*	8.7500	13.8889	7.88
6	10	8.2100	13.000	4.85
7	10	7.7200	11.8000	6.7
8	10	8.2300	10.8000	6.3
9	10(1)*	8.2900	9.6667	4.3
10	10	7.8000	9.5000	7.6
11	10(1)*	5.7400	7.2222	2.15
12	10(1)*	7.3600	11.8889	3.95
13	10(1)*	7.7800	11.2222	6.95
14	10	7.8700	8.6000	8.05
15	10	8.1800	9.7000	5.85
16	12	7.0750	8.5833	4.75
17	10	8.6900	13.4000	7.30
18	9	8.2444	10.2222	8.05
19	10(1)*	8.6600	9.6667	8.5
Mean	190(8)*	7.8342	10.1538	

* Missing data for accuracy scores.

ing. One individual, for example, scored a low 7.2222 on accuracy and a low 5.74 on attitude.

Accuracy was found to be only minimally associated with attitudinal scores $(r(182) = .2482, p < .0001)$. Answering a question correctly and completely was not a good predictor of how well the students in this study perceived they were being treated. Conversely, librarians who project positive images do not necessarily answer questions with the highest accuracy. Minutes spent with students apparently did affect the attitudinal

scores they assigned to the librarians. Librarians who spent 4 or more minutes with students tended to get assigned a higher attitude score than those who spent less time (F = 7.592, p < .00001).

In 36 instances, inappropriate referrals and sources or outright wrong answers were given.

It was thought that perhaps differences might be found between questions of a ready reference nature and those involving more extended research. Therefore, the 58 questions were divided into two groups based on the number of sources needed to find the complete and correct answers and the level of difficulty of the questions (as perceived by the investigators) to test the prediction that when ready reference questions are asked, the accuracy and attitudinal scores will be higher. However, the type of question did not affect either attitudinal scores (t (186) = -.30, p = .768) or accuracy (t (177) = 1.10, p = .271). Therefore, the difficulty of the question did not significantly affect student ratings of accuracy or attitude.

SIGNIFICANCE AND USE OF THE RESULTS

The study was intended as a practical one—to gain information and insights on how reference service at ISU might be improved. While the significance of certain relationships was examined statistically, no formal hypotheses were formulated or tested. This section of the paper, then, deals with the authors' perceptions of the value of the results to ISU and with how these results have been and are being used.

As full faculty, ISU librarians are evaluated each year for the distribution of merit dollars. Three areas of performance are scrutinized: (1) practice of librarianship (the equivalent of teaching performed by general faculty), (2) research and scholarly activity, and (3) service. Librarianship (the most heavily weighted component) may also be the most difficult to evaluate in many instances—especially for public service librarians. In evaluating reference activity, impressionistic anecdotes or testimonials from colleagues often replace more objective data. Teaching faculty have traditionally been subjected to regular student evaluations. In a similar fashion, unobtrusive evaluations, such as that reported here, furnish a comparable examination of reference performance from several perspectives, accuracy and deportment among them. Such evaluations allow the quality and character of reference service to be discussed and evaluated at a level more concrete than opinion, conjecture, or speculation.

In considering the results of this study, a consensus must first be arrived at as to exactly what is an acceptable level of accuracy and of attitude. Is 70% accuracy acceptable? Is 50%? Is an attitude score of 7.8 on a 10-point scale what an institution should be aiming for or should tolerate? What level is unacceptable—7, 6, 5? Is the fact that 15% of the questions are dealt with in less than two minutes significant? That 37% are dealt with in less than four minutes?

In making use of the results the librarians involved should be made thoroughly familiar with the methodology of the project and the instrument used. Once the group recognizes that there very well could be problems in the level of service furnished, ideas on how to address them can be solicited, or presented, and discussed in an informal meeting. On one level, simply *recognizing* that one may be perceived in a certain way by a patron, or that two or three minutes may not be an appropriate amount of time to give all questions, or that one may have developed a tendency over the years to point students in the direction of sources rather than lead them, might be enough to solve the problem. With some librarians, the mere fact that they are reminded of possible problems or weaknesses in their performance may be enough to create a self-correcting situation. However, this will not always be the case, and other options should be explored—for example (1) personal interviews for the librarians falling at the

464   College & Research Libraries   September 1991

low end of the rating scales; (2) use of outside speakers to present a workshop on improving reference service and combating and reducing the effects of burnout; and (3) identification of the types of questions most likely to be dealt with inadequately.

From an unobtrusive study of the type described, improvement in reference service can be addressed at several levels: personal, divisional, and institutional. If warranted, personal conferences with the librarians can be conducted to discuss, for example, undesirable elements of service. This might be a tendency to use inappropriate reference sources, to conduct peripheral business at the reference desk, or to give an undesirable impression of one's approachability, friendliness, or willingness to help. At this personal level, one can simply run through the list of comments made by the surrogate users and discuss the individual questions with the librarians.

On the divisional or institutional level, the collective consciousness relating to reference service can be heightened by broad, nonconfrontational group discussion of patterns detected. Traditional assumptions and platitudes about the excellence of service furnished can be challenged and strengths and weaknesses pointed out. Librarians with an accuracy score below some selected level should be consulted privately. The pattern of time spent on questions may be worth discussion with some librarians (one librarian spent one minute or less on half the questions received and less than three minutes on 80% of them), as would the collection of comments made by observers (about 7 pages for each librarian).

Traditional assumptions and platitudes about the excellence of service furnished can be challenged and strengths and weaknesses pointed out.

The third level for discussion would occur at the divisional level. Here, if the assessment of performance showed real excellence, as it did in some instances, it can be commended and serve as a morale builder. If, on the other hand, undesirable trends were disclosed (e.g., reluctance to handle questions dealing with a certain collection located on the floor), they should be discussed and existing policy regarding them clarified or revised. One unfortunate aspect of providing anonymity in such a project is that, while the identities of the underachievers are protected, so too are the identities of the stars—the librarians whose performance is truly exemplary and who should be used as role models.

After conducting personal interviews, general and divisional meetings, and an in-house developmental institute, the library should implement a similar project, after an appropriate amount of time has passed, to determine what changes, if any, have occurred as a result of the evaluation process.

REFERENCES AND NOTES

1. Terence Crowley, "Half-Right Reference: Is It True?" *RQ* 25:59–68 (Fall 1985).
2. Jo Bell Whitlatch, "Unobtrusive Studies and the Quality of Academic Library Reference Services," *College & Research Libraries* 50:181–94 (Mar. 1989).
3. Duane E. Webster, "Examining the Broader Domain," *Journal of Academic Librarianship* 13:79–80 (May 1987); and Joan C. Durrance, "Reference Success: Does the 55 Percent Rule Tell the Whole Story?" *Library Journal* 114:31–36 (Apr. 15, 1989).
4. Ronald Rowe Powell, "Reference Effectiveness: A Review of Research," *Library and Information Science Research* 6:3–19 (Jan.-Mar. 1984); Crowley, "Half-Right Reference," p.59–68; F. W. Lancaster, *If You Want to Evaluate Your Library . . .* (Champaign: Univ. of Illinois Graduate School of Library and Information Science, 1988); and Peter Hernon and Charles R. McClure, *Unobtrusive Testing and Library Reference Services* (Norwood, N.J.: Ablex, 1987).

5. Rolland E. Stevens and Donald G. Davis, Jr., *Reference Books in the Social Sciences and Humanities*, 4th ed. (Champaign, Ill.: Stipes, 1977); Thomas P. Slavens, *Informational Interviews and Questions* (Metuchen, N.J.: Scarecrow, 1978); Marcia J. Myers and Jassim M. Jirjees, *The Accuracy of Telephone Reference/Information Services in Academic Libraries: Two Studies* (Metuchen, N.J.: Scarecrow, 1983); Charles R. McClure and Peter Hernon, *Improving the Quality of Reference Service for Government Publications* (Chicago: American Library Assn., 1983); and Janine Schmidt, "Reference Performance in College Libraries," *Australian Academic and Research Libraries* 11:87–95 (June 1980).
6. Wyma Jane Hood and Monte James Gittings, *Evaluation of Service at the General Reference Desk, University of Oregon Library* (Eugene: Univ. of Oregon, 1975), ERIC Document Reproduction Service No. ED 110 038.
7. Thomas Childers, *The Effectiveness of Information Service in Public Libraries: Suffolk County. Final Report* (Philadelphia: Drexel Univ., School of Library and Information Science, 1978), passim.
8. Charles A. Bunge, "Factors Related to Reference Question Answering Success: The Development of a Data-Gathering Form," *RQ* 24:482–86 (Summer 1985).