Unobtrusive Studies and the Quality of Academic Library Reference Services

Jo Bell Whitlatch

This article uses empirical data from a recent obtrusive study of reference performance to explore content validity and assumptions regarding unobtrusive studies. Data collected by the author support the contention that improvements are highly desirable before conducting more unobtrusive studies of reference service. The two most important changes concern the development of test questions representing all types of queries and supplementing the correct fill rate with other measures of reference performance.

During the past two decades, the most notable advance in reference services evaluation has been the increased use of unobtrusive observation methodology. Researchers have come to accept unobtrusive studies as a valuable tool for the evaluation of reference services. Reference librarians and library managers, however, do not appear to have integrated the findings from these unobtrusive studies into reference services practice.

Unobtrusive studies of reference services were developed to provide an alternative to user satisfaction surveys. The early, obtrusive and generally global surveys of user satisfaction provided little information useful for improving services. Unfortunately, unobtrusive studies as they are presently employed in the evaluation of reference services also have serious limitations that prevent an adequate assessment of reference services performance.

This paper begins with a comparison of performance measures and methodologies typically employed in unobtrusive and obtrusive studies and then discusses the assumptions underlying unobtrusive studies. A recently published text, *Unobtrusive Testing and Library Reference Services* by Peter Hernon and Charles McClure, provides an excellent review of unobtrusive methodology and practice.[1] To illustrate differences between these two types of studies, I use material from Hernon and McClure and selected findings from an obtrusive study that I have just completed.[2] My findings support the contention that unobtrusive studies, as they are currently conducted, are extremely limited as instruments for the evaluation of reference effectiveness in academic libraries. The paper concludes with a discussion of the changes that are needed in order to develop an improved system of reference evaluation.

SUMMARY OF FINDINGS AND METHODOLOGY

Unobtrusive Studies

In unobtrusive studies of reference services, predetermined test questions are administered to reference librarians. These librarians are unaware that their responses are being assessed. Thus, the effect of being tested should not influence

Jo Bell Whitlatch is Associate Library Director, Access and Bibliographic Services, at San Jose State University, San Jose, California 95192-0028.

the normal behavior of the librarian. Those administering the test questions pose as library users and receive training in how to administer the questions to the unsuspecting librarians. The results indicate that reference staff members answer correctly 50 to 60 percent of the questions posed in this manner.[3]

Thomas Childers remarks that he and Crowley unintentionally initiated a tradition of research and a particular way of conceiving the reference process and applying the unobtrusive method to the evaluation of reference services. Today, those performing unobtrusive studies continue to conceive the reference process and employ the unobtrusive method in roughly the same way as the original investigators.[4]

Typically, unobtrusive studies use the correct answer fill rate, that is to say, the proportion of correct answers to questions, as the measure of reference performance. In their recent study of government documents reference service, Hernon and McClure employed the two types of test questions most commonly used in unobtrusive studies: factual, e.g., requests for the name of an individual or for specific statistical or descriptive information; and bibliographic, e.g., requests for bibliographic citations, information on the availability of a publication in the library or through the Government Publications Sales program, or information on obtaining a Superintendent of Documents classification number.[5] They report that the most frequent reasons for incorrect answers are that the library staff member gave wrong data (96 cases or 64.4 percent), responded "don't know" without referral (30 cases or 20.1 percent) or incorrectly claimed that the library did not own a source (23 cases or 15.4 percent).[6]

Obtrusive Study

This obtrusive study includes 397 reference transactions in five academic libraries in Northern California. Librarians asked users to complete a questionnaire for every fifth reference transaction; librarians also completed a companion questionnaire for every sampled transaction. Matching questionnaires were returned for 257 transactions. Prior to collecting the sample data, librarians from each of the participating libraries met with the researcher to discuss survey procedures. They were made aware of the importance of not biasing the survey results by selecting preferred questions or treating surveyed users differently. Individual librarian confidentiality was guaranteed. Reference departments participated voluntarily in the survey because they wished to obtain an accurate picture of the quality of their services.

The study tests a model of the major variables influencing academic library reference service outcomes. Three measures of reference service performance outcomes were employed: librarian judgments of the value of reference service, user judgments of the value of reference service, and user success or failure in locating needed material. Independent variables used in the study included measures of task uncertainty, time constraints of users and librarians, feedback, and type of reference assistance. Only findings useful in evaluating the role of unobtrusive studies in reference performance are reported in this paper.

Questions from the study were classified into three categories: (1) bibliographic citation for which a correct answer could have been predetermined; (2) questions of fact for which a correct answer could have been predetermined; and (3) other questions, including narrow and broad subject questions; questions concerning evaluations of books, movies, and plays; and questions on how to use reference sources. A small proportion (11.3 percent) of the requests were for specific factual information and 18.0 percent were related to locating specific citations. The majority (70.7 percent) of the queries were requests for locating references on a subject and/or assistance in how to use library reference sources (see table 1).

Results of SPSS cross-tabulations for factual, bibliographic, and subject/instructional questions by user success in locating materials are presented in table 2. The chi-square statistic is significant, indicating that there is a difference between user success in finding material related to

factual and bibliographic queries versus that for other types of queries. For factual and bibliographic queries, greater proportions of users either found what they needed, or nothing that they needed. For subject and instructional queries, a much greater proportion of users found some but not all needed material.

TABLE 1

TYPE OF QUESTION

Type	Number	Percent
Factual	29	11.3
Bibliographic	46	18.0
Subject/Instructional	181	70.7
	256	100.0

Missing (1)

TABLE 2

TYPE OF QUESTION AND USER SUCCESS IN FINDING NEEDED MATERIAL

Materials Available	Question Type Fact.	Bibl.	Subj./Instr.
Yes	78.6%	70.5%	62.6%
Some	10.7	13.6	33.3
None	10.7	15.9	4.1
	100.0%	100.0%	100.0%
	(28)	(44)	(171)

Missing (14)
$X^2 = 16.87$, df = 4, p = .0021

These results are not always directly comparable to unobtrusive findings because in this study some of the material needed to satisfy factual and bibliographic queries was not located because of circulation and collection development problems. In many unobtrusive studies problems with collection development and circulation failures are fairly well controlled through preselection of standard reference works that are likely to be in the library at all times.

ASSUMPTIONS

Accurate Fact Provision as a Key Indicator of Reference Performance

The first assumption is that correct answer fill rate is a key measure of reference service effectiveness.[7] Hernon and McClure have carefully considered some important aspects of the validity of the test questions and measures of accuracy such as face, internal, external, and construct validity. Their questions were judged by librarians and researchers as representative of typical questions encountered at the documents and general reference desk.[8]

However, there has been little discussion of content validity. Content validity is related to the adequacy with which important content has been sampled and the adequacy with which the content has been cast in the form of test items.[9] Therefore if we are really interested in measuring the performance of reference desk service we must ask how well a test represents the main body of reference questions.

Childers roughly estimates that the kind of query that has been addressed through unobtrusive testing to date may represent about one-eighth of the range of reference questions asked.[10] Childers suggests that research findings on part of the process are being taken to represent the whole. The query with a short, factual, unambiguous answer has attracted almost all of the field's attention. The problem with investigating such queries is that in the minds of many of those interested in evaluating reference performance, findings from unobtrusive studies assume unrealistic proportions and come to represent the whole of the reference function. However, there is no empirical literature that links performance of one kind of reference service to performance of another kind.[11]

Evelyn Daniel observes that traditionally fact provision has not been a major service of the library. It became a convenient afterthought to the referral and provision of bibliographic information.[12] Duane Webster suggests that accuracy may not be a key indicator of the overall quality of reference services; users often seem to value convenience and timeliness of information more than accuracy.[13] The findings of the current study support these observations and provide evidence that requests for specific factual information represent a minority of reference queries in academic libraries. This study indicates that the majority of queries are related to broad and narrow subjects or involve requests for instruction in the use of

library reference materials (see table 1).

With such a relatively small percentage of factual queries, librarians get little opportunity to develop on-the-job expertise using a broad range of tools to answer requests for specific factual information. Some factual queries may represent more difficult problems for librarians than subject queries in locating useful information. Success rates for factual queries appear to fall more frequently into the categories of total success or total failure (see table 2). Total failure rates for factual queries may be higher because of the design of bibliographic access systems. Library bibliographic access systems tend to be designed to locate materials by broad subject topic rather than by precise fact. The public catalog still provides the primary access to a library's resources and is normally useful only for locating books. With rare exceptions, the catalog does not provide access to tables of contents or individual chapters in books; neither does it provide access to book indexes, which are most useful for locating factual information.

User demand for factual answering services appears to be relatively low compared to other types of requests for reference assistance. This is particularly true for queries related to academic course work and research.

Results of SPSS cross-tabulations for factual, bibliographic, and subject/instructional questions by purpose of the user are provided in table 3. The chi-square is significant, indicating that there is a difference in proportion among types of queries made for course work, research, and other reasons. The proportion of factual questions asked to meet course work and research needs is much lower than those asked for other reasons, i.e., other job related, personal, and miscellaneous. Of the total number of factual questions in the sample, 42.9 percent are primarily related to course work, 17.9 percent for research and 39.3 percent for other reasons. Therefore, of the 237 questions for which both purpose and type of question are identified, only 17 (7.2 percent) are both factual and closely related to the primary mission of the academic library, support of coursework and research (see table 4).

One must also be realistic by asking if accuracy is a key indicator of reference performance. While unobtrusive test questions have documented, authoritatively correct answers, most real-life questions are not so conveniently documented. Professional education stresses the identification of appropriate sources containing an answer, but not the ability of the librarian to judge the accuracy of that answer. A performance measure that evaluates and identifies a proper source might be a more reasonable test of reference librarian effectiveness.

Patrick Wilson raises serious questions about the ability of reference librarians to determine accuracy in all subject areas. He notes that reference works do not collectively give a single standard answer for the same question; they are, in varying degrees, full of inaccuracies. Further, a standard reference work quickly becomes dated and incomplete.[14]

Wilson concludes that librarians generally work in a world of texts that they take as simply given and cannot claim to evalu-

TABLE 3

TYPE OF QUESTION
AND PURPOSE OF QUESTION

Question Type	Purpose		
	Course Work	Research	Other
Factual	7.5%	10.0%	42.3%
Bibliographic	16.8	26.0	15.4
Subject/Instructional	75.8	64.0	42.3
	100.1%	100.0%	100.0%
	(161)	(50)	(26)

Missing (20)
$X^2 = 28.99$, df = 4, $p < .0005$

TABLE 4

QUESTIONS RELATED AND UNRELATED TO INSTITUTIONAL MISSION

Type of Question	Number	Percent of Sample
Course Work and Research		
Factual	17	7.2
Bibliographic	40	16.9
Subject/Instructional	154	65.0
Other Reasons		
Factual	11	4.6
Bibliographic	4	1.7
Subject/Instructional	11	4.6
	237	100.0

Missing (20)

ate independently.[15] The evaluation of the content of texts requires expertise in the subject matter of the text, which the librarian cannot be expected to have. Librarians are not generally in a position to be able to evaluate the contents of a reference book or to make independent judgments on the correct or incorrect status of answers. It seems, therefore, that our key performance measures ought to be designed to acknowledge more thoroughly the limited judgments librarians are able to make.

Data from this study support Wilson's observations. Librarians in the study reported good subject expertise (1 or 2 on a scale of 7) for only 51.8 percent of the transactions, and users reported the same level of subject expertise only 17.3 percent of the time. This is the nature of general reference desk service, where librarians cannot hope to have in-depth subject competence in all areas for which they are expected to answer questions. One user noted, "The people are helpful, and try to do their best, but *some* of them are not qualified enough." Significant subject familiarity by the librarian was positively associated with user success for subject and instructional questions ($r = .303$, $p = < .0005$) and factual and bibliographic queries ($r = .208$, $p = .040$). User reports of shorter lengths of time spent with the librarian were more strongly associated with user success in locating materials for factual and bibliographic citation queries ($r = .560$, $p. = < .0005$) than for subject and instructional queries ($r = .149$, $p. = .026$).

Hernon and McClure note that when conducting evaluation studies, issues related to the quality of the service must be separated from the value of that service. Ultimately, the value of the service is based upon the degree to which the service meets the information needs of library clientele and facilitates the accomplishment of service objectives.[16]

Libraries and librarians should be judged primarily on whether they provide added value to users. When partial and full success are both considered, users report greater success in finding some or all materials for subject and instructional (95.9 percent) queries than for factual (89.3 percent) and bibliographic (84.1 percent) queries. Only when judged by the more stringent criterion of locating all materials needed did users report the greatest success rates for factual queries (see table 2). Thus, the correct answer fill rate appears to be a useful, but extremely limited, measure of reference performance.

Easier-than-Average Queries

Another assumption underlying many unobtrusive studies of reference service is that the questions used are not difficult to answer. McClure and Hernon suggest that the 55 percent correct answer fill rate is typically computed on questions with an "easier-than-average" difficulty level.[17] Also, in response to a reviewer's query as to whether the degree of difficulty should be used to judge the quality of reference service, Hernon and McClure state that "factual and bibliographic questions are generally recognized as two of the easier types of reference questions."[18]

However, the degree of difficulty of average unobtrusive test questions versus average questions actually asked in academic libraries for factual, bibliographic, and subject/instructional questions has not been carefully studied. For service providers, task difficulty was found to be positively related to task uncertainty.[19] Thus, data on task uncertainty collected for this study provide an opportunity to

explore the differences in types of questions.

"Work flow uncertainties are created by unpredictable client arrival, service, and exit patterns."

Uncertainty can be defined as a situation in which one cannot control or reliably predict all of the variables and relations.[20] Uncertainty is also thought to have an important influence upon delivery of services.[21] Fundamental differences exist between production processes in service and manufacturing industries. These important differences include work flow and task uncertainties peculiar to service operations. Work flow uncertainties are created by unpredictable client arrival, service, and exit patterns. Task uncertainty occurs when there is incomplete knowledge about how to produce a desired outcome. Because the production of service outcomes depends upon interaction between clients and service providers, workers cannot totally rely upon past procedures when providing service to individual clients.

Thus, this obtrusive study includes five uncertainty measures for each question. The five measures are librarian ratings of: (1) the frequency of use of sources used to answer a question, (2) the question as a new type of problem, (3) the similarity of a question to other questions, (4) familiarity with the subject of the question, and (5) familiarity with the information source used to answer a question.

Librarian ratings are compared for factual, bibliographic, and subject/instructional types of queries. Factual and bibliographic citation questions are separated for this analysis because the majority of questions used for unobtrusive studies appear to concern factual rather than bibliographic citations. Thus, this study compares mean ratings of task uncertainty for the factual, bibliographic citation, and subject/instructional types of questions (table 5) and also compares mean ratings of task uncertainty for factual versus all other types of questions (table 6).

Means for frequency of use of sources are significantly different for factual and subject/instructional questions, with sources for factual queries used less frequently (table 5). For three measures of task uncertainty—similarity of questions, librarian subject familiarity, and librarian information source familiarity—means for factual and bibliographic questions are significantly different.

For three measures—frequency of use of sources, similarity of questions, and librarian subject familiarity—there are significant differences between the means for factual and other types of reference queries (table 6). On the average these librarians regard factual questions as somewhat less routine, because they involve the use of somewhat less frequently used sources and are slightly less similar to other types of questions. Librarians also report somewhat less subject familiarity when responding to the factual queries included in this sample.

". . . librarians judge factual questions to be more difficult because answering these questions involves the use of less familiar, less frequently used sources."

Mean ratings of the task uncertainty involved with factual queries indicate that on the average librarians in this study view the uncertainty in the task of answering factual queries as somewhat greater than the uncertainty involved in answering other queries. Therefore, findings in this study suggest that librarians judge factual questions to be more difficult because answering these questions involves the use of less familiar, less frequently used sources.

Represents Real-Life Patrons

Hernon and McClure argue that increased use of unobtrusive testing of refer-

TABLE 5

MEAN RATINGS OF TASK UNCERTAINTY FOR FACTUAL, BIBLIOGRAPHIC, AND SUBJECT/INSTRUCTIONAL QUESTIONS
(Rating Scale: 1 = very great or completely; 7 = very little, very seldom or not at all)

Task Uncertainty	Fact.	Bibl.	Subject/ Instruct.	df	F
Frequency of use of sources	2.79[a]	2.07	2.03[a]	2,249	3.34*
New type of problem	6.00	6.04	5.73	2,252	.40
Similarity of questions	3.38[b]	2.15[b]	2.69	2,253	5.20†
Librarian subject familiarity	3.59[b]	2.48[b]	2.96	2,253	3.52*
Librarian information source familiarity	2.59[c]	1.59[c]	2.25[c]	2,253	5.53†

*p < .05
†p < .01
[a]Means for Factual and Subject/Instructional queries are significantly different at the .05 level (Scheffe test).
[b]Means for Factual and Bibliographic queries are significantly different at the .05 level (Scheffe test).
[c]Means for Factual and Subject/Instructional queries are significantly different from Bibliographic queries (Scheffe test).

TABLE 6

MEAN RATINGS OF TASK UNCERTAINTY FOR FACTUAL VERSUS OTHER TYPES OF QUESTIONS
(Rating Scale: 1 = very great or completely; 7 = very little, very seldom or not at all)

Task Uncertainty	Type of Question		df	F
	Factual	All Others		
Frequency of use of sources	2.79	2.04	1,250	6.67*
New type of problem	6.00	5.79	1,253	.44
Similarity of questions	3.38	2.58	1,254	6.22*
Librarian subject familiarity	3.59	2.86	1,254	4.27*
Librarian information source familiarity	2.59	2.11	1,254	2.83

*p < .05

ence services is necessary because it provides an excellent means to see the library and its services from the viewpoint of the patron.[22] However, unobtrusive testing takes the view that information is a commodity. Catherine Sheldrick Ross notes that if we think of information as a commodity, as existing apart from people or their contexts, then questions and answers can be held to exist independently.[23] Library schools have typically taught and evaluated basic reference courses this way. They give students questions to answer. This approach strengthens the tendency of these future librarians to conceptualize information as a commodity with no reference to its context in the life of the inquirer.

In reality people ask questions to fill gaps in their understanding so that they can use the information as a means to an end. Users are helped to the extent that the answers to their questions help them accomplish something. Librarians who answer questions without knowing anything about the context may provide an answer that is technically correct but is otherwise unhelpful in filling the user's need.[24]

Fred Oser summarizes trends on the basis of a survey of the literature and concludes that there is a large area of situational content that can be of great use to the librarian in conducting an efficient and rapid interview. The type of library in which one is working can lead to highly predictable expectations toward purpose, scope, and level of reference queries.[25] Helen Gothberg also notes the variability in levels of service, which are based to a considerable extent on the type of library. For example, a special library with its limited clientele may find it most expedient to provide the answer. On the other hand, libraries located in educational institutions place a greater emphasis on educating the

user to answer his/her own questions at the basic level of information need; yet there need be nothing minimal about the type of instruction provided.[26]

In this study, a substantial proportion (23.9 percent) of the bibliographic citation questions are identified in the context of the library. That is, the library user had an index or other bibliographic material in hand and wanted the librarian's assistance with locating one or more references and/or understanding the meaning of different parts of the citation. This relatively straightforward, but nonetheless important class of bibliographic citation questions does not appear to be represented in unobtrusive studies.

The unobtrusive model of testing reference service assumes that the librarian is responsible for finding the correct answer. A proxy is hired to conduct the test. Hernon and McClure note that research has shown that many users of academic and public libraries are not aggressive in pressing staff for an answer.[27] They also indicate that, ultimately, the responsibility for ensuring that the patrons' information needs are met belongs to library management.[28]

Proxies neither suggested sources or places where the answer might be obtained nor encouraged referrals. The methodology also makes librarians, not users, responsible for locating and verifying the exact information. Hernon and McClure note that when library personnel referred a proxy to a source, but did not offer to look for the answer, the proxy would pretend to examine the source for a short time and return to the same person for further guidance in use of the source.[29]

But in reality, user behavior may significantly affect reference performance. Wilson notes that in the delivery of reference service, limits are set by the preferences, habits, abilities, and resources of the user. The library cannot supply the user with time or ability. It can supply documents to study, but not the inclination to do so.[30] In this study, librarian judgments of the value of service are significantly related to librarian judgments of user participation in the reference process for factual and bibliographic citation questions ($r = .236$, $p = .022$) but not significantly related for subject and instructional queries ($r = .111$, $p = .073$). User success is positively associated with librarian perceptions of feedback quality both for factual and bibliographic questions ($r = .262$, $p = .014$) and for subject and instructional questions ($r = .155$, $p = .024$). When users let the librarian know how well the question is answered, this feedback is significantly and positively associated with all three reference service outcomes, irrespective of the type of question, again suggesting that the controls placed on proxy behavior in unobtrusive studies may lower the success rate (see table 7).

"... the controls placed on proxy behavior in unobtrusive studies may lower the success rate."

For factual questions, approximately one-third of the users report receiving a direct answer, while other users report receiving assistance in locating the answer for themselves (table 8). Therefore, in these five academic libraries, users report that the librarian accompanied them (but did not refer) or provided a direct answer for 65.5 percent of all factual queries. Librarians did not have the opportunity to verify the complete precise information for the remaining 34.5 percent of factual queries. Consequently, in the real-life provision of factual reference service, academic librarians often do not assume full responsibility for direct answer provision

TABLE 7

CORRELATIONS BETWEEN USER FEEDBACK AND REFERENCE SERVICE OUTCOMES BY TYPE OF QUESTION

Reference Outcome	Feedback on Factual/Bibl.	Feedback on Subject/Instruct.
Librarian judgment of service value	.160	.200†
User judgment of service value	.593‡	.387‡
User success	.247*	.306‡

*p<.05
†p<.01
‡p<.001

TABLE 8

TYPE OF REFERENCE ASSISTANCE FOR FACTUAL/BIBLIOGRAPHIC QUESTIONS

Assistance	Factual
Accompany	31.0%
Refer	10.3
Accompany & refer	24.1
Direct answer	34.5

and verification of accuracy of information.

Instructions to proxies do not encourage them to push the librarian to provide as much assistance as needed. Should so much emphasis be placed on an evaluation method that eliminates many of the contextual clues provided by users seeking to fill a gap in their understanding and to use the information for their own purposes?

An alternative model to use in studying the reference process is one in which the reference transaction and the resulting service outcome are joint products of the effort, skill, and knowledge of both librarian and user. In Gordon Whitaker's model of coproduction of service, service delivery is accomplished through a process of mutual adjustment, whereby both client and service provider mutually consider the client's problem and what each of them can do about it. Citizens influence the content of many public services through direct participation in service delivery. This is particularly the case in services designed to change or empower people directly, such as education and health care. The agent can encourage, suggest options, illustrate techniques, and provide guidance and advice, but cannot bring about the change alone.[31] For libraries functioning as part of the educational process, this may be an appropriate model.

CONCLUSION

John Campbell suggests that given a specific research question, we can ask what specific research method(s) possess the most validity for the independent and dependent variables. It should not be assumed that hard measures are always more valid than soft measures of dependent variables. Campbell comments that the term "hard" seems to refer to dependent variables that consist of countable outcomes. They are objective in the sense that they can be counted by an independent party. Soft measures seem to refer to those based on human judgment or scaling consideration.[32]

The unobtrusive studies use a modified laboratory method and have the strengths and weaknesses commonly found in such methods. Their strengths are that reference librarians cannot introduce bias into the study because they are unaware of it, and a clear standard for correct and incorrect answers to queries can be established. Their weakness is that they do not necessarily represent reality. Figure 1 provides a summary of the differences in treatment of users and questions for unobtrusive versus obtrusive studies of reference performance.

Studies employing unobtrusive methods are often viewed by researchers as more desirable for reference evaluation because of their objective qualities. Unobtrusive studies seem to be more scientific. Hernon and McClure express this view in the following statement: "Basing managerial decisions regarding reference services on perceptions rather than realistic appraisal is a disservice to library clientele and a myopic stance that continues to impede the development of quality reference services."[33]

The popularity of unobtrusive studies appears to be a reaction to early uncritical global surveys of user satisfaction with libraries and the growing awareness that although users appear to be highly satisfied with library service, they do not represent the best critical judgment about the provision of information. Perhaps we have now gone too far in the opposite direction by studying an unrepresentative minority of reference queries.

The method has been allowed to dictate the evaluation criteria and scope by limiting the test to only that portion amenable to hard, objective measures. The scientific method was originally developed in order to more effectively study the physical world around us. Is this method really the primary one to use for the study of how

Unobtrusive Studies	*Obtrusive Studies*
	Users
—are passive receivers of service	—may be passive or active
—have little investment in answers	—have a definite interest in results
—operate without a context in which to place answers	—have a knowledge gap to fill
—provide limited or no clues to librarians	—provide a wide range of contextual clues to librarians
—operate without specific time pressures	—operate under a variety of time pressures
—expect librarians to supply the answers	—normally expect librarians to provide a source for answers
	Questions
—are quantifiable requests for specific information	—are general requests for subject information
—are relevant to a limited task	—are relevant to a broad set of tasks

FIGURE 1
Differences between Unobtrusive and Obtrusive Reference Studies

people behave in organizational settings?

Unobtrusive studies tell us only a little about the quality of reference services. Given the weaknesses in unobtrusive studies and the problems with content validity, can more representative studies be designed using the unobtrusive methodology? Unobtrusive methods remain a valuable alternative methodology for countering some of the known weaknesses in field studies using obtrusive methodologies.

But unobtrusive studies would tell us much more about the quality of reference services if the assumptions and the methodology were modified substantially. Test questions need to represent all facets of reference. Unobtrusive studies must develop more representative reference questions. Research findings from the present obtrusive study demonstrate that it is feasible to involve reference departments in the evaluation process and collect summaries of questions and answers. Collecting and compiling such information also enhances the ability to analyze the entire spectrum of queries people bring to libraries.

Questions representing the entire body of reference could be selected from questions sampled in the field, and expert peer review could be used to supply appropriate sources. Unobtrusive observations should also be used more extensively to collect information on referrals and service orientation (the helpfulness, responsiveness, and interest of the librarian in the user's problem).

Wilson observes that librarians can claim to be adept at locating texts and what these texts say about each other and the external world.[34] Therefore, the correct referral to appropriate sources would be a more appropriate measure than accuracy. Even correct referral to sources is not without its problems as a performance measure. Sandra Naiman notes that other professions can and do agree that there is a basic core of information and or skills that members must posses. Yet she has never met a group of librarians who were willing to reach a consensus on the indispensable reference sources.[35]

Reference performance measures used in obtrusive studies could also be modified for use in unobtrusive studies. Charles Bunge and Marjorie Murfin have established a stringent criterion for patron-perceived fill rate.[36] In order to count as a totally successful question, users must report that they located just what was wanted and that they were completely satisfied with the information or materials found or suggested. The average success rate for thirty-one libraries participating in their study was 55.81 percent for all types of reference questions and 46.7 percent for factual reference

questions. Murfin and Gary Gugelchuk note that the unrealistically high ratings found in previous studies of reference performance may be due in great part to the use of inadequate instruments and methods to study a complex phenomenon.[37]

In the present study of five academic libraries, 66 percent of users report finding what they needed. This exceeds the average success rates generally found in unobtrusive studies. The majority (75 percent) also report that they are very highly satisfied (1 or 2 on a scale of 7). For this study, adopting a more stringent criterion similar to that used by Bunge and Murfin would result in a total success rate of 57 percent for users who found what they needed and also indicated that they were very highly satisfied. Therefore, use of these or similar measures in unobtrusive studies would definitely permit researchers to include the more common subject and instructional types of questions with more precise measures than global indications of satisfaction.

With the evidence of poor performance on certain types of factual and bibliographic questions that librarians are asked, academic librarians might be well advised to be more critical in interpreting the text to the user, particularly when they cannot claim expertise in the subject area. Library schools might consider providing more extensive education for librarians in interpretation and verification of answers in standard factual sources.

Academic librarians responding to factual queries in areas for which they lacked the expertise to make an independent evaluation would then say to the user, "Here is a source that might help you find an answer to your question" rather than, "Here is the answer to your question." When problems of authority arise and neither librarian nor the user have the subject expertise to make a judgment, academic librarians would do well to refer users to teachers with expert knowledge.

The results of the unobtrusive studies also provide convincing evidence that many librarians need to be more critical in assessing their expertise when responding to factual queries. Harold Wilensky notes that one standard highly adhered to and accepted in established professions is awareness of the limited competence of your own specialty within the profession and readiness to refer clients to a more competent colleague.[38] By closely adhering to this professional standard, reference librarians could improve the quality of their factual answering services by referring users to expert sources rather than providing an answer of uncertain quality. For academic libraries with a government publications department, specialized factual answering services might be combined with the duties of documents librarians. Librarians in smaller libraries who have difficulty developing such expertise in the staff might try to maintain sources for expert referral and carefully identify those sources for which they are prepared to provide in-depth factual question answering services.

Because of the weaknesses of both unobtrusive and obtrusive studies, more sophisticated methods must be developed for evaluating reference service performance. When reference evaluation methods are able to provide a more comprehensive picture of the quality of reference service, reference librarians and managers will be more interested in using the results to modify reference service practice. A major advance in improving reference evaluation will be expanding the scope of the predetermined test questions now used in unobtrusive studies and developing additional measures of effectiveness to supplement correct answer fill rate.

While we are waiting for more sophisticated studies that use a greater variety of evaluation methods, we can modify our professional service practices by more critically examining our competence to answer certain types of factual queries. For most queries, academic librarians might do best to focus on evaluating their competency in source referral, both internal and external. Finally, we should consider how often we ask users whether their questions were answered fully and whether they have found what they wanted.

REFERENCES

1. Peter Hernon and Charles R. McClure, *Unobtrusive Testing and Library Reference Services* (Norwood, N.J.: Ablex, 1987).
2. Jo Bell Whitlatch, "Client/Service Provider Perceptions of Reference Service Outcomes in Academic Libraries: Effects of Feedback and Uncertainty" (Ph.D. diss., University of California, Berkeley, 1987).
3. Peter Hernon and Charles R. McClure, "Unobtrusive Testing: The 55 Percent Rule," *Library Journal* 111:37–41 (1986).
4. Thomas Childers, "The Quality of Reference: Still Moot After 20 Years," *Journal of Academic Librarianship* 13:73–74 (1987).
5. Hernon and McClure, "Unobtrusive Testing," p.39.
6. Ibid., p.48.
7. Hernon and McClure, "Unobtrusive Testing," p.24.
8. Ibid., p.41.
9. Jum C. Nunnally, *Psychometric Theory* (New York: McGraw Hill, 1978), p.93.
10. Childers, "The Quality of Reference."
11. Ibid.
12. Evelyn H. Daniel, "The Effects of Identity, Attitude and Priority," *Journal of Academic Librarianship* 13:76–78 (1987).
13. Duane E. Webster, "Examining the Broader Domain," *Journal of Academic Librarianship* 13:79–80 (1987).
14. Patrick Wilson, *Second-Hand Knowledge: An Inquiry into Cognitive Authority* (Westport, Conn.: Greenwood, 1983), p.173–92.
15. Ibid., p.192.
16. Hernon and McClure, "Unobtrusive Testing," p.128.
17. Ibid., p.144.
18. Ibid., p.167.
19. Richard J. Butler, "User Satisfaction with a Service: An Approach from Power and Task Characteristics," *Journal of Management Studies* 17:1–18 (1980).
20. James D. Thompson, *Organizations in Action* (New York: McGraw Hill, 1967).
21. Peter K. Mills and Dennis J. Moberg, "Perspectives on the Technology of Service Operations," *Academy of Management Review* 7:467–78 (1982).
22. Hernon and McClure, "Unobtrusive Testing," p.104–105.
23. Catherine Sheldrick Ross, "How to Find Out What People Really Want to Know," *The Reference Librarian* 16:19–30 (1986).
24. Ibid.
25. Fred Oser, "Referens Simplex or the Mysteries of Reference Interviewing Revealed," *The Reference Librarian* 16:53–78 (1986).
26. Helen M. Gothberg, "The Beginnings," *The Reference Librarian* 16:7–18 (1986).
27. Hernon and McClure, "Unobtrusive Testing," p.167.
28. Ibid., p.105.
29. Ibid., p.39–40.
30. Patrick Wilson, *Public Knowledge, Private Ignorance: Toward a Library and Information Policy* (Westport, Conn.: Greenwood, 1977), p.122.
31. Gordon P. Whitaker, "Coproduction: Citizen Participation in Service Delivery," *Public Administration Review* 40:240–46 (1980).
32. John N. Campbell, "Labs, Fields, and Straw Issues," in *Generalizing from Laboratory to Field Settings*, ed. Edwin A. Locke (Lexington, Mass.: Lexington Books, 1987), p.269–79.
33. Hernon and McClure, "Unobtrusive Testing," p.164.
34. Wilson, *Second-Hand Knowledge*, p.192.
35. Sandra M. Naiman, "The Unexamined Interview Is Not Worth Having," *The Reference Librarian* 16:31–46 (1986).
36. Charles A. Bunge and Marjorie E. Murfin, "Reference Questions—Data From the Field," *RQ* 27:15–18 (1987).
37. Marjorie E. Murfin and Gary M. Gugelchuk, "Development and Testing of a Reference Transaction Assessment Instrument," *College and Research Libraries* 48:314–38 (1987).
38. Harold Wilensky, "The Professionalization of Everyone?" *American Journal of Sociology* 70:137–58 (1964).

APPENDIX A: SCALES FOR VARIABLES INCLUDED IN THE AUTHOR'S OBTRUSIVE STUDY

Reference Service Performance Outcomes

1. Librarian Judgments of the Value of Reference Service (Five Questions, loading on one factor, a = .89)
   Please check the space:_____: that best describes how you think the user viewed the quality of service received. (seven-point scale: outstanding—terrible)
   Relevance of Information Provided. (two seven-point scales: useful—useless; relevant—irrelevant)
   Amount of Information Provided. (two seven-point scales: sufficient—insufficient; reasonable—unreasonable)
2. User Judgments of the Value of Reference Service (Six Questions, loading on one factor, a = .84)
   Check the space:_____: that best describes the general quality of service you received. (seven-point scale: outstanding—terrible)
   Indicate how satisfied you are on the following scale. (seven-point scale: satisfactory—unsatisfactory)
   Relevance of Information Provided. (two seven-point scales: useful—useless; relevant—irrelevant)
   Amount of Information Provided. (two seven-point scales: sufficient—insufficient; reasonable—unreasonable)
3. User Success (One Question)
   Were you able to locate the materials you needed? (choices were: yes; no; some but not all; and other (please explain))

Task Uncertainty

(Five Questions, loading on one factor, a = .80)
To what extent were the sources you suggested to this user materials you frequently consult in providing reference service? (seven-point scale: very great extent—very little extent)
To what extent did you see answering this reference question as a new type of problem: (seven-point scale: very great extent—very little extent)
How often do you answer this question or questions that are very similar? (seven-point scale: very often—very seldom)
How familiar were you with the subject(s) involved in the reference question? (seven-point scale: completely—not at all)
Were you already familiar with the information resources most likely to contain the answer to this reference question from previous knowledge or experience? (seven-point scale: completely—not at all)

Time User Spent with Librarian

(One Question)
How long did you spend with the reference librarian? (choices were: 0-2 minutes; 3-5 minutes; 6-15 minutes; over 15 minutes)

User Feedback

(One Question)
To what degree did you inform the library staff member whether or not your question was answered? (seven point scale: completely—not at all)

Librarian Perception of Quality of Communication

(Four Questions, loading on one factor, a = .72)

Communication with the user was: (two seven-point scales: very easy—very difficult; pleasant—unpleasant)
Did the user give you sufficient information to answer his/her question? (seven-point scale: sufficient—insufficient)
How explicit was the user's question? (seven-point scale: very explicit—not at all explicit)

Librarian Judgments of User Participation

(Two Questions, loading on one factor, a = .72)

To what extent did the user provide you with feedback? (seven-point scale: very great extent—very little extent)

How active a role did the user play in resolving his/her information need? (seven-point scale: very active—not very active)

Type of Reference Assistance

(One Question)

How did the reference librarian assist you? (the user selected one of five choices: (1) by accompanying you to sources to help find the answers; (2) by referring you to sources to find the answer on your own; (3) by accompanying you to some sources and referring you to other sources; (4) by directly giving you the answer to your question; or (5) other (please explain))

NOTE: The abbreviation "a" refers to Cronbach's alpha, a measure of internal reliability for the variable.

The library program is basically concerned with teaching and learning, with adapting the library to instructional needs, and with improving student work and achievement.

—Frank A. Lundy, January 1951

William Miller, in The Book Industry, *quotes G. P. Bret, Jr., president of Macmillan Company, as saying in March 1948 that "all (book) costs have increased since 1940 between 60 and 70 percent."*

—Elizabeth C. Seely, October 1951