V. REFERENCE EVALUATION

Reference Evaluation:  
An Overview

Patricia Hults

SUMMARY. The current interest in reference evaluation has its roots in the unobtrusive testing begun twenty years ago. Evaluation before that was centered on criteria of quantity rather than quality. When unobtrusive testing began reporting accuracy rates hovering in the 50% range, reference services responded by exploring a variety of techniques to measure quality of service including unobtrusive testing, patron surveys, and peer review. Methods to improve reference service, such as training workshops, peer coaching, and changes in organizational climate are also beginning to be critically examined.

HISTORICAL PERSPECTIVE

The issue of reference evaluation is a relatively new phenomena, particularly when reference performance is considered. Murfin and Wynar identified under 40 articles concerning reference evaluation published between 1920 and 1960 (Murfin and Wynar 188-213; Altman 170). Until the late sixties virtually all assessment of reference service, other than patron...
surveys, involved quantitative measurements: how many questions were answered in which category from what type of patron at what time of day. Much effort was spent in activities such as defining reference vs. directional and hammering out the intricacies of grid statistical analysis. That emphasis on data gathering was a reflection of the same emphasis evident in the broader field of evaluation, particularly governmental programs assessment. Much of the research being done at that time in performance measurement was centered on data gathering instrument refinement and statistical analysis (Cronin 6). Although current reference evaluation efforts focus on performance and qualitative measurements, the statistical information gathered was important and allowed librarians to begin to compare and evaluate the scope of their reference activities.

During this time patron surveys were another method of evaluation frequently used by libraries. However, most of these surveys suffered from the same perplexing problem. According to Rothstein "A number of studies have attempted to ascertain the opinion held by reference clientele regarding the service received, and the results could hardly be bettered by paying for testimonials" (Rothstein 464; Altman 175). Patrons were so happy with any service they seemed unable to make distinctions of quality. In the sixties the field of performance measurement was further stimulated by the requirement that government programs contain an evaluative component. Evaluation research became more sophisticated and began to go further than simple data gathering (Cronin 6). The subject of library evaluation also began to elicit increased interest and show increasing sophistication. In 1967 Crowley conducted one of the first unobtrusive testings of reference performance (Crowley; Weech and Goldher 306).

**UNOBTRUSIVE TESTING**

Crowley continued his research and in 1971 Childers and Crowley published their landmark work in which they reported that reference staff correctly answered about 50-55% of the questions posed (39-51, 139). Similar figures have repeatedly appeared in subsequent studies done across the U.S., in England by House (Altman 175), and Finland by Koivunen (James 97). Hernon and McClure have furthered the research with consistent and similar results hovering around the 55% accuracy level.

The basic design of unobtrusive testing is simple. Proxies are trained to pose as library patrons and ask factual reference questions. These ques-
tions are normally designed so that they can be answered by sources held by the libraries being studied. The librarian's answer is rated as to its "correctness." Some studies differentiated the level of correctness. Other studies rated additional factors, such as level of the librarian's education, size of the collection, or librarian's communication skills, and tried to come up with predictive factors for the accuracy rate. The results have been mixed. In general, there appears to be an insignificant correlation between collection size, library budget, demand, physical setting and correct fill rate (Gers and Seward; Crews). Professional librarians do seem to score higher than non-MLS degree holders, although not all studies report this (Crews 341-42). A study done in Maryland concluded that the highest predictive factor of success is the individual librarian's behavior such as reference question negotiation skill, librarian interest and comfort with the question, and perhaps most importantly, follow up (Gers and Seward 33).

The library community responded to these unobtrusive studies first with shock, then with denial. The common response is "yes, maybe, but not in my library." Charles McClure speaks of hearing that statement again and again, often from librarians who work in the very libraries he has tested and found right there at the magic 55% mark. As study after study confirmed the 55% figure, librarians then began poking holes in the methodology of unobtrusive testing. An article by Bill Bailey is typical of this response. He argues that the data was flawed because of test design errors including lack of third party observation of the interaction and controls on timing of questions and busyness at the desk. He also felt that "the point is that surreptitious observation eventually will uncover flaws in even a paragon of professionalism. Hernon and McClure could have tailed the brightest reference librarians until they finally gave out wrong answers" (281). The real point is that Hernon and McClure didn't do that. They tested librarians in their normal work situations and librarians failed.

Others have questioned the validity of the questions posed and felt that measuring a small aspect of the total picture of reference work was an inaccurate reflection of the quality of reference work (Whitlatch).

All of these responses beg the question. No method of evaluation measures the totality of a job or is free of design error. What the library community really needs to address are the questions McClure and Hernon have left us with: is a 55% accuracy rate acceptable; if not, what priority do libraries place on improving that rate; what is the cause of the rate and what is the cure? (Hernon and McClure, Library 70).

Certainly accuracy of information is not the only criteria of quality reference service, though it seems the baseline to work from. We are all
familiar with the librarian who, though she may be accurate, has such a stern "bed-side manner" as to be unapproachable to the average patron. This obviously is not quality service. However, accuracy and "bedside manner" are not necessarily independent variables. The Maryland study indicates the highest predictive characteristics for reference accuracy have to do with those "bed-side manners," i.e., reference negotiation skills, interest in the question, etc. (Gers and Seward 33). This was confirmed by the improvement in accuracy experienced by the librarians in a program developed and instituted by Maryland's Division of Library Development. The program trained librarians in those positive reference communication behaviors. Follow up studies reported dramatic improvement in correct answer rates (Arthur 368). This makes intuitive sense. Librarians skilled at reference negotiation and other forms of communications with patrons have the best chance of understanding the question and thus providing accurate information. Reference evaluation does need to look at more than one facet of the reference interaction to accurately measure the service provided. The important characteristics, accuracy, reference interview negotiation skills, approachability, etc., are synergistic and work as a whole to determine quality of service.

Hernon and McClure have begun to look at what institutional climates best support quality reference service. There is some research backing their hypothesis that an institution that is innovative, supportive, democratically governed, and with high morale will provide better reference service ("Unobtrusive Testing and Library..." 107; "Unobtrusive Testing and the Role..." 74). Lowenthal reports preliminary correlations between emotional well-being (morale) and job performance in a study done in public libraries. Degree of disaffection from patron, depersonalization, anxiety, stress, etc. all had significant negative impact on reference performance (385-392).

**CURRENT DEVELOPMENTS**

Some libraries are responding to the challenge of these studies by beginning to develop methods to routinely assess reference service. A few libraries have conducted their own unobtrusive studies. The Cumberland County Public Library and Information Center took the bold step of using board members as their test proxies. Their risk paid off in increased board involvement and commitment to library excellence. Interestingly, as part of the study the proxies were asked if they were satisfied with the service they received. Although the overall accuracy rate was 74% and the prox-
ies knew when the answer was incorrect, the proxies were satisfied 90% of the time, another example of the halo effect libraries seem to have (Hansel).

The Fairfax County Public Library also conducted an unobtrusive test of services. Because of lack of funding to hire outside evaluators, they developed a cooperative effort with a neighboring library system. Staff of both libraries were involved in the study plan, design, and execution. It was felt that this involvement was important to acceptance of the results by insuring that the staff viewed the test as fair and appropriate. Both libraries used their testing to identify needed policy manuals, appropriate training, and in some cases, staffing reorganization (Rodger and Goodwin).

Few libraries have the resources to conduct annual unobtrusive tests. Other methods have been developed to incorporate regular reference review. Linda Olson has broken reference service into four components and suggested methods of assessment for each. Instead of unobtrusive testing to measure the librarian’s ability to provide factual and policy information, she suggests a 4 hour test administered to each reference librarian. For measurement of the staff’s ability to provide instruction she suggests carefully designed user surveys.

One interesting method just developing in reference service is peer review. The Ramsey County Public Library developed a comprehensive system that included outside consultants to provide training in confrontation, nonverbal expression, and problem identification. The evaluation began with an extensive self-evaluation form that was duplicated and distributed to all reference staff. Facilitated meetings then used the peer review format to explore competency strengths, common factors affecting performance, reference objectives and duties. The entire format of the evaluation served as a training workshop for improved reference service ("Peer"). Other libraries with peer review programs include Louisiana State University Libraries, Auraria Library in Denver, the libraries at the University of Arizona, Appalachian State University in North Carolina and Bowling Green State University (Kleiner 353).

DEVELOPING A REFERENCE EVALUATION PROGRAM

Before you can begin to develop a reference evaluation program you must decide what you are going to evaluate; program quality, individual performance, or both. This choice will drive the choice of your tools of measurement. For example, unobtrusive testing is more appropriate for
program evaluation than for individual assessment. The unique character-
istics of your institution will also determine which evaluation tools will
work best for you. Departments with large numbers of staff have much
different logistical problems that those with small numbers. If your staff
normally works alone, as is common with small staffs, peer review be-
comes more difficult to organize.

Early staff involvement is important regardless of the focus of your
evaluation. Involving the reference staff in the process of developing
criteria and evaluation methods will promote a democratic operation of the
reference department and increase morale by insuring a much higher level
of acceptance and enthusiasm for the end product. Early involvement in
the development will promote staff acceptance because their philosophies
and concerns will be incorporated into the structure. The process itself
should also improve reference quality by ensuring that each reference
librarian understands the goals of the reference department, understands
what characteristics are considered good or bad reference practice, by
promoting awareness and sensitivity to the issue of reference quality, and
by creating a shared vision of what reference service should be.

A useful step is an examination of the goals of the reference depart-
ment. You can't measure a service without a clear idea of what it is you
want it to do. Bunge’s article is a thorough explanation of the process of
creating goals. The process is a multi-layered effort with the goals at
every level being fueled by the broader institutional goals. Adams and
Judd also focus on goal setting and do an excellent job of breaking down
goals into specific, measurable performances.

The danger of this approach is the tendency to create goals of quantity
rather than quality. Ellison and Lazeration recommend a reference evalua-
tion method based on management by objectives whose main focus is on
countable activities peripheral to the main function of reference service.
An example is an objective that the librarian will contact X number of
new faculty per semester. This method does not evaluate the actual refer-
ence performance.

The development of criteria to measure quality is the next step. Several
checklists already exist that pinpoint specific behaviors of reference librar-
ians such as does the librarian smile, make eye contact, take the patron to
the source, acknowledge a waiting patron (Gers and Seward 34; Adams
and Judd 137-43)? The reference staff can use these as jumping-off points
to develop their own list appropriate for evaluation. These could be used
by supervisors, peers, or for self-evaluation. They also can be used as
standards and training tools to raise librarian awareness of the importance
of these traits. The development of the criteria should help clarify the
program evaluation than for individual assessment. The unique characteristics of your institution will also determine which evaluation tools will work best for you. Departments with large numbers of staff have much different logistical problems that those with small numbers. If your staff normally works alone, as is common with small staffs, peer review becomes more difficult to organize.

Early staff involvement is important regardless of the focus of your evaluation. Involving the reference staff in the process of developing criteria and evaluation methods will promote a democratic operation of the reference department and increase morale by insuring a much higher level of acceptance and enthusiasm for the end product. Early involvement in the development will promote staff acceptance because their philosophies and concerns will be incorporated into the structure. The process itself should also improve reference quality by ensuring that each reference librarian understands the goals of the reference department, understands what characteristics are considered good or bad reference practice, by promoting awareness and sensitivity to the issue of reference quality, and by creating a shared vision of what reference service should be.

A useful step is an examination of the goals of the reference department. You can't measure a service without a clear idea of what it is you want it to do. Bunge's article is a thorough explanation of the process of creating goals. The process is a multi-layered effort with the goals at every level being fueled by the broader institutional goals. Adams and Judd also focus on goal setting and do an excellent job of breaking down goals into specific, measurable performances.

The danger of this approach is the tendency to create goals of quantity rather than quality. Ellison and Lazeration recommend a reference evaluation method based on management by objectives whose main focus is on countable activities peripheral to the main function of reference service. An example is an objective that the librarian will contact X number of new faculty per semester. This method does not evaluate the actual reference performance.

The development of criteria to measure quality is the next step. Several checklists already exist that pinpoint specific behaviors of reference librarians such as does the librarian smile, make eye contact, take the patron to the source, acknowledge a waiting patron (Gers and Seward 34; Adams and Judd 137-43)? The reference staff can use these as jumping-off points to develop their own list appropriate for evaluation. These could be used by supervisors, peers, or for self-evaluation. They also can be used as standards and training tools to raise librarian awareness of the importance of these traits. The development of the criteria should help clarify the
librarian's understanding of appropriate reference technique and promote consistency in the application of these behaviors. Once criteria have been agreed upon the method(s) of measuring those criteria must be tackled. The decision either to measure the department or the individual will determine the direction you take. Some tools appropriate for departmental evaluation are unobtrusive testing and patron survey. Hernon and McClure's articles provide good background reading for unobtrusive testing and Hansel and Rodger and Goodwin's articles outline approaches that can be taken on a relatively small scale.

Patron surveys need to be developed very carefully to be useful. These surveys need to be designed to measure specific components of instruction—did the librarian take you to the source, did the librarian explain in a logical manner how to use the source, did the librarian follow up later by asking if you had found what you needed? Hopefully these types of questions would gather more useful evaluation of instruction. Linda Olson's article is particularly helpful for survey development as well as Murfin and Gugelchuk.

If you are examining individual performances several methods have been used. The traditional one of supervisory rating may be the method most prone to subjectivity and bias. Some organizational structures require supervisory evaluation for reappointment, tenure, or promotion, but combining it with peer or self-evaluation will improve chances of accurately evaluating an individual. Articles useful when considering peer evaluation are Kleiner and "Peer Evaluation of Reference Librarians in a Public Library."

You can, of course, combine methods. The broader the information gathered about your reference service the more accurate the picture will be. The process of developing a reference evaluation program is not a trivial one. Considerable staff time and commitment are needed for the process to succeed. As the program is being developed and tried out, evaluation and revision are necessary as weaknesses and strengths of particular methods in your particular institution are identified.

**AFTER EVALUATION**

The literature stresses again and again the need to tie evaluation to some results. "If plans are not made to act on the results of an evaluation process, one might save the time and money involved . . ." (Green 168). One obvious response to evaluation is to provide training for areas of measured weakness. There is a whole body of literature concerning staff
training, but of particular interest is some research done by McClure and Hernon.

They examined the effect of one workshop on reference accuracy rate. A library which holds a government document collection was unobtrusively pretested with questions that could be answered with basic materials from that collection. The staff then attended a 4 hour workshop on government document sources designed to increase their awareness and skill in using the government document material. The library was then unobtrusively tested again with the disheartening results that the accuracy rate had dropped.

The design of the workshop was typical of 1/2 day workshops librarians attend all the time. It probably was more carefully run than many. McClure and Hernon felt that even if it did contain serious design flaws (which they doubted), sheer exposure to the material should have raised awareness of that source of information. They conclude that awareness does not necessarily translate into skill of use and that one-shot workshops are not the answer to improving reference accuracy ("Unobtrusive Testing and Library. . ." 78-103).

An interesting program that seems to address this dilemma was developed at Temple University Libraries. They have had a long standing tradition of staff training, but became uneasy with the issue of transferability of the knowledge from workshop to work site. Drawing on the work done by the Maryland State Department of Education, Division of Library Development and Services, they instituted regular peer coaching for their reference staff. They concentrated on communication skills such as question negotiation and positive non-verbal behavior. They report "(1) greater clarification of the reference process for all staff involved in coaching; (2) increased recognition of positive communication behaviors, both through observation and feedback; (3) increased self-awareness of individual communication style and desk behavior; and (4) increased reinforcement of positive desk behaviors" (Arthur 372). It would have been interesting if they had tested accuracy rates before and after the training to see if they experienced improvements similar to those reported in Maryland.

CONCLUSIONS

Reference departments have slowly begun examining the quality of service they provide. Much work has been done in the last ten years, but much remains to be done, particularly in the assessment of training and other methods to improve reference service. Very few studies have been
done that carefully examine the success of particular techniques of training, the impact of organizational structure on reference work, or the effects of morale, etc. on staff performance.

The process of developing and implementing an evaluation system is not a trivial one. It requires significant staff time and commitment. The implications of effective evaluation, particularly tied with effective staff training, are far reaching for our profession. If we can accurately measure how well we provide that service we say we provide and then improve service with training that works, we will go a long way in our quest for excellence in reference service.

REFERENCES


