Assessing Library Services

A Practical Guide for the Nonexpert

Lisa R. Horowitz

The idea of assessing library services can intimidate even the most seasoned librarian. This article is a straightforward introduction to assessment for the nonexpert, written for reference and public service managers and coordinators who want to analyze data about existing services or incorporate assessment into plans for a new service. It addresses how to approach assessment, how to make it more useful, and how it can improve a service over time. Those who are new to assessment will find this article a helpful starting point for service evaluation; those who are experienced in assessment can share this article with nonexpert staff to explain the basics of assessment, dispel anxieties, and generate support for assessment projects.

Background in Assessment

Many books and articles have been written that contribute to knowledge in the library field of how to do assessment. There are a few that have much more information than this article offers and yet offer introductory information for the nonexpert learning how to do assessment. A major contribution to the field is the *Reference Assessment Manual,* published by the American Library Association (ALA) in 1995. This comprehensive volume of articles about reference assessment aims to "provide practicing librarians, reference managers, and researchers with access to a wide range of evaluation instruments useful in assessing reference service effectiveness."[1] The book comprises three major parts: (1) a set of essays that describe the appropriate research processes and the best tools to answer certain kinds of questions, (2) an extensive annotated bibliography of articles about reference assessment and performance evaluation of reference librarians, and (3) a summary of many of the instruments described in the bibliography, most of which are included on a (sadly now outdated) 3½" floppy disk.

As that book is well more than a decade old, several authors have worked to bring the information in it up to date. Jo Bell Whitlatch's book, *Evaluating Reference Services: A Practical Guide,* updates the *Reference Assessment Manual,* comprehensively including articles from 1995 through the book's publication date, again with annotations. In her book, she also includes a valuable, in-depth description of how to assess reference services.[2] It is an excellent introduc-

tion to evaluating reference services, using simple language and practical suggestions, including sections such as "What is a Survey, and Why Would I Want To Conduct One?" and "Ethical Issues" regarding observation techniques. More recently, Eric Novotny's book of articles specifically addresses how to assess reference in a digital age.[3] The articles cover a broad range of reference modes (chat, e-mail, online tutorials, etc.) and a number of instruments are included as well. Although not necessarily written for the nonexpert, it does update the work of the *Reference Assessment Manual* and *Evaluating Reference Services.*

Reference is only one of many services that libraries assess, and there are a number of books that more broadly address service assessment. A recommended introduction to the evaluation process as a whole is Joseph Matthews' *The Evaluation and Measurement of Library Services.*[4] The chapters describe what evaluation is and its purposes, and what tools to use and why. There are several individual chapters devoted to specific library services as well. A number of other books have been written that provide broader knowledge of how to do research and conduct data analysis, but those move out of the scope of the audience for this article.[5] Others have focused on service quality. Peter Hernon and Ellen Altman describe how to assess service quality with a focus on library users in *Assessing Service Quality: Satisfying the Expectations of Library Customers.*[6] Individual chapters explain how to define performance standards, and when to use certain kinds of tools. Darlene Weingand's book, *Customer Service Excellence: A Concise Guide for Librarians,* uses a libraries-specific approach to the business concept of customer service excellence, with a focus on public libraries.[7]

After the turn of the twenty-first century, the need for assessment data became greater as library resources became more and more restricted. New literature that showcases different ways to assess in order to produce data with more impact were produced. Roswitha Poll and Phillip Payne explain that because resources within institutions and communities have become more and more scarce, libraries must find ways to prove their contribution to learning,

Lisa R. Horowitz (lisah@mit.edu) is Coordinator of Central Reference Services at Massachusetts Institute of Technology Libraries in Cambridge.

teaching, and research quantifiably.[8] They do so by concentrating on impact measures for libraries and information services. Larry Nash White provides a case for assessment of intangibles in libraries, as the world has moved toward a focus that is more "service and information oriented than production oriented." He stresses that "library goodness" could be considered one of these intangibles.[9] Peggy Maki, former director of assessment at the American Association for Higher Education, combines an assessment focus on student outcomes with a description for how to plan assessment for that purpose.[10] Steve Hiller and Jim Self, two well-known library assessment experts, describe in one article how data gathered from assessment can help with planning and decision making.[11] That article includes a comprehensive literature review of the use of data in libraries and offers contemporary examples of libraries that have used data successfully for planning and management.

In 2002, a committee of ALA's Reference and User Services Association (RUSA) was tasked to create guidelines for assessment of reference. In the end, there were too many variables to set definitive guidelines, but the committee did develop a toolkit introduced in 2007 that explains in detail what kinds of analysis are best used with which assessment tools. The toolkit points to a variety of tools used by various libraries, and gives an extensive bibliography of references.[12] This toolkit can be a valuable resource for the nonexpert. What none of these articles or books offers is a concise description of what the nonexpert needs to know in order to begin an assessment process. This article attempts to explain the basic premises that assessment requires and how to move towards an assessment of a service having defined a scope for the project that is within budget, works with the staff available, and serves a valuable purpose. The rest of the article has four parts. The first section describes general principles upon which all assessment efforts should be founded. The second section explains how to plan for assessment within the context of the resources and staff time available. The third section explains how to implement these principles and planning, and the fourth section explains post-assessment evaluation—how to assess the effectiveness of your assessment project, and how to move forward.[13]

SECTION I: General Principles

Poll, a frequent contributor of articles on performance measurement (especially for the International Federation of Library Associations), defines "performance measurement" as the

> collection of statistical and other data describing the performance of the library, and the analysis of these data in order to evaluate performance. Or, in other words: Comparing what a library is doing

(performance) with what it is meant to do (mission) and wants to achieve (goals).[14]

Library staff and administrators want to assess services for a variety of reasons: to justify positions or resources, describe patterns of use and demand for services, determine appropriate service hours or staffing, for peer comparison, to improve services, evaluate new services, etc. Most important in the recent literature is assessing the impact of the library and its services on its constituencies.[15] The word "assessment" is used to describe all the various strategies used to satisfy these myriad needs.

Although assessment can be initiated for any number of reasons with questions that can be answered with variety of tools, all assessment strategies should meet common standards to ensure that they are accurate, will lead to improvements, and are appropriate to their specific purpose and institution. By adhering to the following seven basic tenets of good assessment, results will be richer and more likely to be trusted by administration.[16]

The Seven Basic Tenets of Good Assessment

1. Devise a set of performance standards based on the values of your library.

A standard, according to *The Oxford American Dictionary of Current English,* is "an object or quality or measure serving as a basis or example or principle to which others conform or should conform or by which others are judged."[17] Performance standards in this context are those targets, outcomes, or metrics that are desired to indicate that a service is successful. As Whitlach explains, "typically you need to assess against some agreed upon performance standards. You develop goals for service and measurable objectives, and then determine whether the objectives are met"[18] by comparing your actual performance to the standards that you have set.

The performance standards that you employ will be defined by the values of your library. Your institution or unit may already have its own rules about service (e.g., "We respond to any e-mail question by 5 p.m. the next business day.") that you can use as your standard, and you can also rely on or modify other standards for reference and information service (see a list of possibilities in the appendix at the end of this article). When you devise performance standards, consider your goals and what will make your service a success within the context of the values held by your library. Hernon and Altman describe a method that leads from values to standards. They suggest that staff look at a list of "service quality expectations" and identify first-, second-, and third-level priorities among them:

> Such a list of priorities serves as a reminder that the intent is to identify those expectations that staff believe are most essential to meet and to lay

the foundation both for expanding the list over time and for deciding where to set [standards].[19]

The set of priorities identified by staff becomes the foundation for performance standards.

Self explains how, at the University of Virginia, where a "balanced scorecard"[20] method of assessment is used, values lead to standards: "By limiting the number of scorecard metrics [items that indicate success within the specific standard], it forces us to decide what is important, and to identify those numbers that truly make a difference."[21] These values will also influence what you assess and why, which leads to the next principle.

2. Determine what questions you want answered.
Many people begin by looking at data that already exists, such as data reported for accreditation or generated by the catalog, and then try to determine what that data tells them about their services. Although there is value to analyzing preexisting data, preexisting data can only answer very specific questions; as a result, limiting oneself to mining preexisting data severely restricts the range of questions that can be answered. The most relevant assessment instead begins by asking questions generated by your performance standards. If a performance standard requires that you respond to any question by 5 p.m. the next day, your questions might include, *Are we meeting our question response-time goal? How often do we miss that goal? Can that be improved? Are our users satisfied with our response time?* As questions based on specific standards are generated, it will be found that there are many possible directions to explore.

3. Make sure to use the most appropriate data-gathering tool for your questions.
Some people begin assessment by choosing a particular data-gathering tool, even before their questions have been determined. For example, they begin by deciding that a focus group or a survey would be a good idea because they need input from library users, and they feel these are the logical ways to gather it. However, this is an ill-considered practice that can restrict you to answering only those questions that can be answered by that tool. The list of possible questions to ask is unlimited if it is considered before choosing a tool. A best practice is to let the questions determine what data is needed, and then allow data needs to point toward the most appropriate tools.[22] There are a number of tools that can be used, and different tools may offer quantitative or qualitative feedback. Qualitative feedback is non-numerical and may be generated by surveys, interviews, focus groups, or even observation. Quantitative feedback is often generated through unobtrusive observation, such as gate counts or circulation numbers, but is sometimes imposed on qualitative feedback, such as where a scale from 1 to 5 is assigned to something that is qualitative. This is described in more detail in tenet 4.

Not all assessment tools serve every purpose, and both quantitative and qualitative methods have value. As Danny Wallace and Connie Van Fleet explain,

> the specific methodology for the evaluation project devolves from the perspective established for the evaluation effort. Certain methodologies lend themselves to certain perspectives. Focus groups, for instance, are very useful in determining patron perceptions, are of limited use in determining patterns of use, and are of extremely limited value in assessing efficiency or effectiveness.[23]

Focus groups are guided, moderated discussions on a topic familiar or meaningful in some way to the participants involved. Because of their format, focus groups and one-on-one interviews are not usually representative enough to be generalized to a large population. This kind of qualitative feedback can best be used to solicit ideas, such as for new library initiatives, or to corroborate and expand on ideas learned from a larger-scale survey. Surveys, in turn, are useful for gaining perceptual and opinion feedback of a large population, but are less useful for confirming factual information. For example, if users respond that they visit the library once a month, it means that users "reported" they visited the library once a month. But to learn about actual patterns of use or to corroborate facts, direct observation of intermediated transactions or indirect examinations of electronic transactions, gate counts, etc., are more useful. Poll and Payne explain the value of combining qualitative and quantitative methods:

> The results of qualitative methods will of course have a subjective bias; they show the "perceived outcome." They should therefore be compared with results of quantitative methods or with statistics of library use in order to validate the results. But the "anecdotal evidence" will be invaluable in reporting to the public and the institution, as it serves to make statistics understandable and believable.[24]

For those interested, an assessment glossary has been compiled that offers complete definitions of these methodologies, which can help determine the best use of each tool.[25] Tools are also described in more depth later.

4. Define plans for data use.
Regardless of the tools used, ensure that all data collected is used in a contextually appropriate way. Additionally, make sure the interpretation of data will suggest actions that can improve services or fulfill the needs of assessment identified above. Communication of the use of data and its results needs to be clear as well. Staff can get frustrated when they must gather data simply because it is easy to gather, but they see no outcomes from the data collection.

Users can become frustrated if they are asked questions, but no action results. Maki writes, "Assessment is certain to fail if an institution does not develop channels that communicate assessment interpretations and proposed changes to its centers of institutional decision-making, planning and budgeting."[26] It is equally important to share that information with users as well. This demonstrates to all the stakeholders that their participation in assessment activities has had value, and creates an atmosphere where change and improvement are expected and accepted.

How data is used also has a bearing on the kind of data that needs to be collected. Decisions will need to be made about whether qualitative or quantitative data will answer your question, and, if quantitative, what measurement level is needed.[27] This level describes which scale is used in rating questions. One of four common types of scales is used, depending on whether the numbers in the scale are truly continuous. In library studies, a numerical rating is often assigned to satisfaction, success, or an evaluation of an instructor. The importance of the measurement level is that with continuous numbers, statistical measures are legitimate. Using certain statistical and arithmetical functions on non-continuous numbers leads to invalid or questionable results. The least numerical scale is the *nominal scale,* in which numbers are merely used as named categories; for example, the library call numbers or a phone number. The numbers do not offer a measurement, and cannot be used in a statistical way. On the other end of the spectrum is the *ratio scale,* the most truly continuous scale, which can be used in any statistical calculation. It has a true zero point and can be divided into infinitesimal pieces which remain meaningful. This scale includes numbers such as people's ages, or lengths of time. The *interval scale* is similar to a ratio scale, in that the differences between the values are fixed and can be divided, but the value of zero is arbitrary, such as with temperature scales or calendar dates. The zero point does not actually represent a lack of the feature being measured. This means that certain kinds of statistical work can be done to these numbers, but not all.

Library assessment often involves the *ordinal scale.* Ordinal scales define rank–how satisfied users are or who is the best teacher–but do not define a particular distance from one level to another. In library studies, there is often an ordinal scale that shows a satisfaction rating, a success rating, or an evaluation of an instructor. For example, a write-up explaining that library satisfaction scored "3.1" on a scale from 1 to 4 is ordinal. The problem is that the distances from 1 to 2 or 3 to 4 are not necessarily the same (*I am very satisfied* vs. *I am satisfied* vs. *I am dissatisfied*, etc.). Even the number ranks themselves likely mean something different from one person to the next. So although the ordinal scale can be valuable, using it arithmetically to get a mean, for example, can be misleading.[28] Consider the implications of the measurement scale used when creating assessment tools and applying them in interpreting results.

5. Triangulate the data.
Corroborate results by collecting data through several different methods of assessment, ideally three. As RUSA's guide states, "the more sources of data, the better your analyses will be . . . [T]riangulation increases the validity of your analyses and results."[29] Valid results flow from multiple data streams, rather than relying on a single assessment tool. This may involve compiling the results of several quantitative analyses, or could be a combination of qualitative and quantitative data. For example, in a study of virtual reference at the Massachusetts Institute of Technology (MIT), results of two user surveys alongside chat and e-mail transcripts were analyzed, along with activity levels and return rates for usage of the chat service, and usability studies.[30] Taken together, the results offered a rich variety of data indicating how MIT libraries' virtual reference services could be improved. In another example, when deciding how to envision MIT reference services for the future, users answered a survey and focus groups were held to support or extend the information received from the survey, which was more comprehensive. Staff members were surveyed, and all the data examined together and used to write a "reference vision" that could be used into the future.[31]

6. Endeavor to maintain statistical validity.
Maintaining statistical validity (occurring when variables truly do measure what they represent) and reliability (occurring when the tools used can be trusted to measure it consistently) is probably the hardest thing for the nonexpert to do. However, the most reliable results are statistically valid, will stand up to investigation by administrators and outside reviewers, and can be corroborated. Gathering numerical data with the appropriate measurement level is a good start.[32] There are, additionally, some tools to assist with statistical validity, such as random number generators and sample size calculators.[33] RUSA's guide offers suggestions for analytical tools that are best in certain contexts. If ensuring that results are statistically valid is a concern, a statistician can often be found at an institution or locally. Other sources are RUSA's Directory of Peer Consultants and Speakers (https://cs.ala.org/speakers), or conferring with someone from ALA's Office of Research and Statistics (www.ala.org/ala/ors/researchstatistics.htm).

7. Pretest data collection instruments.
Make sure that questions in each instrument are understood by those who will be answering them, and that they measure what is desired. Ensure the data gathered by the tool actually answer initial questions. Use a small group representing those who will be taking the survey, and watch them take the test. Consider having them answer the questions out loud. Be certain that subjects interpret the questions as intended, and do not find your questions ambiguous. Look at the "results" as though they are real, to be sure the results obtained answer the questions you asked. When you analyze the results, can you interpret

them? Sometimes, as much as your questions seem to make sense, the data gathered in response to the questions does not actually add anything to what you already know. Testing the tools helps to avoid that problem. By incorporating these seven basic principles into your thinking about and planning for assessment, you will find that assessment is easier to manage and less intimidating.

SECTION II: Planning Assessment

Assessment involves a number of steps. The first, and perhaps most important, is to take time to plan. Writing out a plan helps to manage the scope of the assessment, and makes it less intimidating. By writing or at least outlining a plan, how much is done at any time can be controlled. Doing so will also help reassure staff that incorporating assessment into their daily workflow will not be burdensome. An assessment plan will be most effective if it meets the following three criteria: (1) keeping a focus on actions for improvement or change, (2) maintaining a pace that makes the plan manageable, and (2) avoiding or addressing potential obstacles that can hinder the assessment process.

Focus on Actions

Assessment is most welcome when it results in positive changes or improvements, whether to users, workflow, use of resources, or determination about how to proceed with a new service. When planning assessment, keep in mind anticipated positive changes or improvements. This will help limit the scope of the assessment, as activities which are less likely to produce any sort of change can be eliminated or minimized. Efforts can then be focused on activities that will result in beneficial changes.

Make it Manageable

Assessment becomes unmanageable when its scope reaches beyond the control of the person implementing it. Make it manageable by limiting the scope where possible.

- Assessment takes time for all the people concerned—the plan writers, the creators of the tools, those who input data, those who analyze the data. Make sure the assessment plan can be managed within the amount of staff time available. Make the project smaller if staff are not able to take on a large project. That does not mean that the assessment will necessarily be less valuable. The impact of a smaller assessment project may pave the way for administrative support for additional staff time for assessment in the future.[34]
- Another way to ensure that time is available for assessment is to tie assessment to a project that is already planned. When planning to redesign reference space, incorporate assessment into that project. If starting a new training initiative, describe how to assess the success of the initiative. By making sure that assessment is not separate but is part of projects already planned, there is a logical space for assessment in the workload, and time becomes easier to manage.
- Plan the assessment over a cycle of several years, rather than considering it a one-time project. This enables it to be ongoing, it allows for new questions over time and it limits expectations of ever having a "complete" assessment at any one point, which is certainly not manageable.[35]
- Make the plan modular. Having an assessment plan with different components that can be handled simultaneously or at unrelated times works well within a cycle of assessment. Tie various standards to certain years (for example, get user feedback on services every three to five years, but focus on workflow improvement in other years). Alternatively, focus on certain units or services within each year, without overloading staff with assessment activity. This modularity also keeps assessment manageable by reducing the expectation of what can be achieved at any one time, while nonetheless illustrating that in the long term, a variety of assessment projects (and related improvements!) can take place.
- Do not overwork your human "resources" devoted to assessment. There is a pool of survey subjects, focus group participants, willing interviewees, and library staff. Try not overload them with too many assessment requests at any time, so as not to use up their willingness to participate. Make sure that the plan does not dictate the use of these "resources" many times within a short period.[36]
- Finally, assign responsibility for assessment. Assessment is less likely to stall if a particular person or group champions it. This champion should represent the stakeholders in the assessment process, people who will be affected both by the assessment process itself, and by the changes that may take place based on the results. This champion can then advocate for funding support and staff commitment to the assessment process, and can maintain history and continuity over time, as assessment needs change.

Overcoming Obstacles

The previous section on managing assessment hints at obstacles that may hinder assessment, with the goal of handling them during the planning process. This section highlights possible obstacles and suggests how to prevent them from hampering that process. There are three kinds of obstacles that can hamper good assessment. First, administrative support comprises buy-in, funding, and allocated staff time or positions. If administrators do not believe that assessment can result in positive changes and improvements,

start smaller. Select an area that requires less support from them, and show them how valuable assessment can be. Consider writing the assessment plan in conjunction with stakeholders to encourage buy-in. Maki advises communicating regularly with stakeholders to promote support for assessment:

> If an institution aims to sustain its assessment efforts to improve continually the quality of education, it needs to develop channels of communication whereby it shares interpretations of students' results and incorporates recommended changes into its budgeting, decision making, and strategic planning as these processes will likely need to respond to and support proposed changes.[37]

Assessment can require tools, supplies such as postage or incentives, or even simply training for staff. Ideally, library administration will support the assessment initiative and provide some funding when needed.[38] However, lack of funding does not mean that assessment cannot be done. If there is not an assessment budget, look for tools that are free or inexpensive (see RUSA's guide for ideas). Consider allotting more staff time for a homegrown product, if possible. Again, when shown that assessment produces results, it may be possible to get future funding for something with a larger scope. In the written plan or outline, make it clear what can be done within the funding available currently, and explain what might be done with more funding.

The second type of obstacle is internal. As mentioned before, it is crucial that staff support the assessment and believe that it will result in positive change. Otherwise, it will not be supported, and may be resented. Staff who are resentful or who do not believe in the process may forget or refuse to provide data or input when it requires manual action. Conversely, they may anticipate an outcome that negatively impacts them, and therefore may provide inaccurate data in an attempt to skew the results. The best way to combat this type of obstacle is to share the assessment plan and what it means with the staff involved. Make sure that they understand the likely positive impacts that their cooperation will produce.

The third type of obstacle is external–the users themselves. Respondents can hinder assessment through low response rates or through sham responses. Response rate is important because low response rates from a representative sample can reduce the validity of generalizing the results to the whole population.[39] (Response rates from a non-representative sample are less critical in that the results cannot be applied to the general population anyway.) Response rates can be raised through incentives for participation, support and encouragement from a critical stakeholder whose opinion makes a difference to the respondents, or through clear benefits in how the results will be used. Sham responses are much less likely but can actually occur as the result of incentives. If the incentive is very

good (e.g., every responder will receive $10), there may be respondents who answer randomly simply to obtain the incentive. Generally, this is something that cannot be totally avoided except by limiting incentives.[40]

SECTION III: Implementing the Assessment Plan

Up to this point, the basic tenets for doing assessment and the planning required to make it effective have been described. The following is a suggested list of steps that lead to assessment that is manageable and "actionable." It relies on the standards created, which in turn depend on the values of the institution. This leads to a set of measurable objectives, which then point to the appropriate tools to obtain those measurements. Finally, there are suggestions for where to go for input on how to analyze and interpret the data that is gathered.[41]

Ask Questions Based on Standards

As described above, the assessment will rely upon a set of standards. In MIT Libraries' Public Services department, for instance, the following were developed based on standards and guidelines from the library literature (see the appendix) as well as from the libraries' own goals:

> 1. Staff provide consistent and excellent service in all interactions.
> 2. Service is both available and delivered in modes appropriate to user needs.
> 3. Users get accurate, appropriate and timely service.
> 4. Users are aware of and understand our services and policies.
> 5. Staff are empowered to be successful in their roles.
> 6. There is effective, efficient and balanced use of resources.[42]

The standards that are listed above are not actually measurable. They suggest, in vague terms, what the ideal service would provide. The University of Virginia Library has similar standards. For example, it includes a standard, "Provide excellent service to users of the University of Virginia Library."[43] Tenet number 2 from the "General Principles" section provides the next step to get to something measurable: Ask questions for which answers are desired. Look at one or all of your standards, and brainstorm questions that will tell whether or not the standard has been met. An example will make this clearer.

At MIT in 2007, public services administrators chose to focus assessment on the standard, "Users are aware of and understand our services and policies." Many committees

and units analyzed that standard from their own point of view. The Research and Instruction Support Group (RISG) brainstormed questions related to research and instruction, such as, "Do people know we offer research assistance in a variety of formats?" and "How do users find out about our research and instruction services?" The group brainstormed a total of almost forty questions. At the same time, the Client Focus Group brainstormed questions related to outreach. In each case, the standard tied the questions together, so that all units were considering users' awareness and understanding of our services.

However, generating the questions is only the first step. Be sure that the project is manageable by reducing the number of questions that will be addressed. It is possible to combine questions into categories and again, going back to the stated values, determine which are the most important to analyze at the moment. Prioritize questions which are most pertinent now, whether for administrators or which lead to actions that can improve the service. RISG generated forty questions, but in the end chose to focus on only a single one, given the time available for this project and the staff available to work on it.

Define Measurable Objectives for Each Question

A measurable objective is just that: an objective, goal, target, that can be measured. Look at the top priority questions generated in the last step, and then reword them into something measurable.

Moving from the questions to something measurable can be the most challenging part of assessment because it means assigning numbers to the institution's values. Especially if you have no idea how well your service is currently doing, it is hard to assign what seems like an arbitrary number to an objective to determine how well you are doing, but it is important to do so. Hernon and Altman agree: "Because there are no universally accepted standards or norms for performance on any aspect of library processes, functions, or service, senior staff in each library have to decide on the acceptable level of performance."[44]

There are several ways to accomplish this. The first is to define numbers that are desirable as well as reasonable. For example, if a standard already exists—let's say that staff e-mail response turnaround is by 5 p.m. the next business day—a measurable objective that states, "Eighty percent of e-mail questions get a complete answer or significant starting help by 5 p.m. the next day, while the other twenty percent get some sort of response in that same time period" can be developed. Those numbers seem completely arbitrary without context. The context of the institution—anecdotal evidence or other data—will guide what is desirable and reasonable.

At the University of Virginia (UVa), standards are translated into what they call metrics which are the specifically measurable way to view the standard. UVa developed two metrics for the standard example above ("Provide excellent service to users of the University of Virginia Library"): "Overall rating in student and faculty surveys [given by the institution]" and "Customer service rating in student and faculty surveys [given by the library]." Each of these metrics then has two targets, which are numbers defined by the staff that have been determined to be desirable and reasonable. The first target for the first metric, "At least 4.25 out of 5.0 from each of the major user groups: undergraduate students, graduate students, humanities faculty, social science faculty, and science faculty," is their most desirable target, while their second target, "A score of at least 4.0 from each of the major user groups," is reasonable and still desirable.[45] If they meet these targets, they feel that they have been successful in this particular area, while if they do not, they have room for improvement.

The second way to develop measurable objectives is more conservative, but can lead to improvements without first assigning a seemingly arbitrary number. This involves answering the questions through data that already exists (or newly generated data) to provide a starting point from which to base future improvement. For example, the question that RISG decided to answer was, "How aware are students, faculty and library service desk staff of our subject guides?" Rather than deciding that a certain percentage of users should indicate awareness of the subject pages (thus imposing an "arbitrary" number based on our values), the group decided to look at a variety of existing data to see whether users did use and understand them. Gathering a variety of data about the subject pages helps determine whether to improve the pages, advertise them better, or discontinue investing resources in them.

If the second method is chosen, the idea is to gather data for a baseline starting point, but the next time assessing the same service, a number that is less arbitrary is applied. For example, after RISG gathered its assessment information in 2007, finding that 20 percent of users were aware of these tools, it set a measurable objective of "50 percent awareness overall (undergrads, grad students, faculty) within three years." The first is a benchmark, while the second is a new measurable objective to strive for.

Once a list of measurable objectives is developed, it can be trimmed to make it manageable. Each objective may have one or several methods that could be used to determine its success. It is ideal to triangulate, as indicated in the principles above, using more than one method to corroborate a set of results. A small assessment project would therefore need to focus on only one or perhaps two measurable objectives. If there is more staff time allotted to assessment, or more projects occurring in which it is logical to incorporate assessment, it may be possible to manage more measurable objectives. No matter how big the project is, the process has now begun: "Setting targets and intended results ensures that the library engages in regular assessment, consistently uses the same questions or statements, and commits the resources necessary to meet its goals."[46]

Design Appropriate Tools

Once defined measurable objectives for each question are identified, the appropriate tool can be applied. The kinds of information needed will determine the best tool for the job.

For example, in the MIT Libraries, RISG evaluated the use of subject guides by looking at the results of a previous library survey question which asked about users' use and awareness of subject guides. Additionally, they compared the number of hits to the subject guide pages versus other libraries' webpages, to see how much the subject guides were used in comparison to how much effort was being put into them. They also examined which of MIT's non-library webpages linked to the subject guides, to get a sense of how often faculty and students advocate the libraries' subject guides. These latter two methods were both unobtrusive observation techniques. For some additional qualitative input, staff were surveyed for anecdotal information about their interactions with users and their use of subject guides, and RISG analyzed a staff self-appraisal that had asked some questions about staff knowledge and use of subject guides. Other than the anecdotal evidence, this was all data that already existed in some format, but by pulling it all together, RISG was able to evaluate both quantitative and qualitative evidence of the use and awareness of the subject guides as well as their value to users.

The Assessment Glossary[47] can assist in explaining the various tools and what they can be used for, as does the RUSA Guide. The UVa balanced scorecard, where the tools are described following the metrics and targets, can also be consulted. In addition, numerous articles and books explain how to use a particular tool or describe an assessment that used a specific tool. For example, there are many excellent books written about how to conduct focus groups, often for commercial settings. One article that is especially useful was written by a consultant for nonprofit organization development, Judy Sharken Simon. It is concise and explains very simply the important points to keep in mind when conducting focus groups. Because its focus is on nonprofits, it has an understanding of the value of a focus group that differs from a traditional commercial approach.[48] Hernon and Altman's book also has a helpful chapter describing how to work with focus groups.[49]

For an explanation of surveys, there is a useful chapter in *Assessing Service Quality*.[50] An especially practical article by the Rand Corporation discusses how to conduct surveys via the Web.[51] It describes the pros and cons of offering a survey in electronic form, explaining various issues related to design and implementation. It is intended for businesses, but much can be gleaned from its uncomplicated description of concerns.

Observation, as explained previously, describes data that is gathered electronically, such as gate counts or circulation statistics. In library terminology, though, observation is often used to describe personal observation of intermedi-ated transactions, whether carried out when the participants are aware (as when a supervisor automatically reviews chat transcripts or watches interactions at a service desk), or through unobtrusive evaluations, for example, when performed by "mystery shoppers." The mystery shopper idea, which comes from business, is when people are asked (or paid) to present questions to a service point, whether planned questions or impromptu, and report certain indicators back to an assessment team. Andrew Hubbertz wrote a rather scathing rejection of this kind of observation that offers thoughtful ideas.[52] His article describes and gives solid evidence showing how unobtrusive observations of actual interactions are often neither legitimate nor needed forms of evaluation. The measurable objectives combined with an understanding of the various kinds of tools available will point toward the most appropriate tools for the job.

Final Steps in Implementation

One key step in implementing the plan is analyzing and interpreting the data. That cannot be described adequately in an article such as this. A good introduction to the variety of possible statistical analysis tools is Hernon and Altman's chapter on "Interpreting Findings to Improve Customer Service."[53] Another set of pointers to statistical tools is in the RUSA guide, and experts can be consulted for assistance, especially if extensive statistical analysis is required.

Many librarians do choose to interpret quantitative findings in qualitative ways rather than using traditional statistical tools. Keeping in mind the tenet to maintain statistical validity, this is possible, as long as caveats related to sampling or other biases in the way the tools are used and/or interpreted are reported. Alternatively, using quantitative data to validate qualitative information can be useful as well. Library directors need to be able to emphasize the "story" behind the numbers to make their case for the value of the library.[54] Once results are analyzed, they can be applied to make improvements to the service or workflow, to justify use of resources, etc. It is critical to communicate with the stakeholders involved. Hernon and Altman point out that organizational values again come to the fore as reported items are highlighted.[55]

SECTION IV: Final Analysis for Success—Assessing the Assessment

After analyzing and interpreting the results, and implementing improvements or actions based on those results, the assessment continues. It is important to assess the assessment process and products regularly. The following are questions that should be asked about the assessment process to help define future assessment:

- Was the process supported by the staff involved? If not, reconsider the process that led to the assessment effort. Involve more staff in the design of the assessment, or in the analysis and implementation of results.
- Were the results shared with users? Did users support and respect the process and the results? It is always valuable for users to understand the work that goes into improving libraries, and the more that is explained and shared with them about the process and resulting actions, the more they will support continued assessment efforts.
- Was the service actually improved? Did the assessment lead to positive action? If not, reconsider the questions asked. Maybe the tool used did not answer that question and therefore led to results that could not be applied. Maybe the tool was less effective than it could have been, and care should be taken to revamp this tool if used again.
- Are the standards still applicable? At UVa, the standards, metrics and targets are reconsidered each year. As priorities and values change, standards may also change.

At MIT libraries, the time came to assess the assessment plan described previously.[56] Several parts of the plan's implementation did not work well. For example, a cycle was imposed in which each standard was examined in a particular year, as described previously. This led to the creation of assessment projects in order to meet the need of the standard, rather than assessment being part of projects already deemed priorities. Second, the standards themselves were reexamined, because it was found that they combined factors that might be better assessed as separate entities. Fortunately, it was also found that the assessment plan had begun to create a sense that assessment was part of the workflow. All new public services projects included an assessment plan, defining how it would be determined that the project was successful. Staff were asking when to switch focus onto a new standard, because there was something else that they thought would be useful to measure but it was not part of the current cycle.

Reassessing your assessment process provides meaning and context to the assessment that you do, and allows it to become a continuous process. When assessment is a regular part of everyone's daily workflow, continuous improvement is a goal that can be obtained.

Conclusion

Assessment of library services does not need to be intimidating, nor does it need to tie up a huge number of resources. If done well, it can lead to positive improvements and changes in services that strengthen a library's relationship to its community. By following basic principles for assessment, planning the assessment to limit its scope, implement-

ing a plan that focuses on positive change, and assessing the process regularly, a library can comfortably incorporate assessment into its regular workflow.

References and Notes

1. American Library Association and Reference and User Services Association Management of User Services Section (MOUSS) Evaluation of Reference and Adult Services Committee, *The Reference Assessment Manual* (Ann Arbor, Mich.: Pierian Pr., 1995).
2. Jo Bell Whitlatch, *Evaluating Reference Services : A Practical Guide* (Chicago: ALA, 2000).
3. Eric Novotny, ed., *Assessing Reference and User Services in a Digital Age,* (Binghamton, N.Y.: The Haworth Pr., 2006).
4. Joseph R. Matthews, *The Evaluation and Measurement of Library Services,* (Westport, Conn.: Libraries Unlimited, 2007).
5. Ronald R. Powell and Lynn Silipigni Connaway, *Basic Research Methods for Librarians,* 4th ed. (Westport, Conn.: Libraries Unlimited, 2004); G. E. Gorman, Peter Clayton, Sydney J. Shep, and Adela Clayton, *Qualitative Research for the Information Professional: A Practical Handbook,* 2nd ed. (London: Facet, 2005). For example, Powell and Connaway's *Basic Research Methods for Librarians,* and Gorman and Clayton's *Qualitative Research for the Information Professional* were recommended to me for that purpose. However, their focus is much less on a nonexpert attempting to do evaluation to improve services than on a librarian trying to do authentic research to improve the profession or for publication purposes.
6. Peter Hernon and Ellen Altman, *Assessing Service Quality: Satisfying the Expectations of Library Customers* (Chicago: ALA, 1998).
7. Darlene E. Weingand, *Customer Service Excellence : A Concise Guide For Librarians* (Chicago: ALA, 1997).
8 Roswitha Poll and Philip Payne, "Impact Measures for Libraries and Information Services," *Library Hi Tech* 24, no. 4 (2006): 547–62.
9. Larry Nash White, "Unseen Measures: The Need To Account for Intangibles," *The Bottom Line: Managing Library Finances* 20, no. 2 (2007): 77–84.
10. Peggy L. Maki, "Developing an Assessment Plan to Learn about Student Learning," *The Journal of Academic Librarianship* 28, no. 1 (2002): 8–13.
11. Steve Hiller and James Self, "From Measurement to Management: Using Data Wisely for Planning and Decision-Making," *Library Trends* 53, no. 1 (2004): 129–55.
12. RUSA, RSS Evaluation of Reference and User Services Committee, *Measuring and Assessing Reference Services and Resources: A Guide,* www.ala.org/ala/mgrps/divs/rusa/sections/rss/rsssection/rsscomm/evaluationofref/measrefguide.cfm (accessed Mar. 22, 2009).
13. Although I am not a scholar in assessment or a statistician,

for almost twelve years I have been working on assessment projects for the MIT Libraries and for the American Library Association's Reference and User Services Association (RUSA). I have been involved in assessing in-person and virtual reference services using surveys, focus groups, and staff forums, as well as detailed analyses of reference transaction statistics, e-mail and chat transcripts, and time logs. Within RUSA, I chaired the Evaluation of Reference and User Services Committee in 2001–02 and was the initiator as well as contributor to a project to produce recommendations for measuring and assessing reference services. In 2007, I led the final process for producing a modern definition of reference for RUSA (http://ala.org/ala/mgrps/divs/rusa/resources/guidelines/definitionsreference.cfm, accessed July 19, 2009). I wrote this article to provide the kind of information that would have helped me in my earlier assessment work.

14. Roswitha Poll, "Measuring Quality: International Guidelines for Performance Measurement in Academic Libraries," *IFLA Publication* 76 (1996): 16.

15. See for examples: Maki, "Developing an Assessment Plan"; Poll and Payne, "Impact Measures"; White, "Unseen Measures"; and Susan J. Beck, "Making Informed Decisions: The Implications of Assessment," *Proceedings of the Eleventh National Conference of the Association of Research Libraries* (Charlotte, N.C.: April 10–13, 2003), www.ala.org/ala/mgrps/divs/acrl/events/pdf/beck.PDF (accessed Mar. 22, 2009).

16. These tenets reflect knowledge that I have accumulated over time. They have been captured in slightly different form in the RUSA Guide, to which I was a contributor.

17. *Oxford Reference Online*, s.v. "standard," www.oxfordreference.com/views/ENTRYhtml?subview=Main&entry=t21.e29962.

18. The idea of basing assessment on guidelines and standards is commonplace. This quote is from personal correspondence with Jo Bell Whitlatch, a well-known scholar in reference assessment, on July 1, 2005. Hernon and Altman said similarly, "Any measure, including customer-related ones, should be linked to performance targets and outcomes, which monitor, and are intended to improve, the quality of service over time," *Assessing Service Quality,* 72.

19. Hernon and Altman, *Assessing Service Quality,* 113.

20. *Oxford Reference Online*, s.v. "balanced scorecard," *Oxford Reference Online,* www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t162.e55.

21. Jim Self, "Using Data to Make Choices: The Balanced Scorecard at the University of Virginia Library," *ARL: A Bimonthly Newsletter of Research Library Issues and Actions,* 230/231 (2003): 2. See also current scorecard metrics at www.lib.virginia.edu/bsc/ (accessed Mar. 22, 2009).

22. RUSA, *Measuring and Assessing Reference Services and Resources: A Guide,* Section 2.0.

23. Danny P. Wallace and Connie Jean Van Fleet, *Library Evaluation: A Casebook and Can-Do Guide* (Englewood, Colo.: Libraries Unlimited, 2001): 7.

24. Poll and Payne, "Impact Measures," 551.

25. Lisa R. Horowitz, *Assessing Library Services: A Practical Guide for the Non-Expert Glossary,* 2009. Access: http://web.mit.edu/~lisah/www/professional/Glossary%20-%20Jan%202009.pdf (accessed Mar. 22, 2009).

26. Maki, "Developing an Assessment Plan," 13.

27. "Measurement level" is defined as **"**the type of measurement scale used for generating scores, which determines the permissible transformations that can be performed on the scores without changing their relationship to the attributes being measured," from *A Dictionary of Psychology,* ed. Andrew M. Colman (Oxford Univ. Pr.: 2006); *Oxford Reference Online,* www.oxfordreference.com/views/ENTRYhtml?subview=Main&entry=t87.e4905 (accessed Mar. 22, 2009)

28. Horowitz, *Glossary,* offers full definitions of the different scales used for quantitative data and the implications for them, as well as other terminology used in assessment.

29. RUSA, *Measuring and Assessing Reference Services and Resources: A Guide,* section 2.0.

30. Lisa R. Horowitz, Pat Flanagan, and Deb Helman, "The Viability of Live Online Reference: An Assessment," *portal: Libraries and the Academy* 5, no. 2 (2005): 239–58.

31. MIT Libraries, "Reference Vision Project Report," 2002, http://libstaff.mit.edu/refcomm/RefVision/documents/refvision_psmg.pdf (accessed Mar. 22, 2009).

32. For full definitions, see Horowitz, *Glossary.*

33. One example of a sample size calculator is offered by Creative Research Systems, at www.surveysystem.com/sscalc.htm (accessed Mar. 22, 2009). This also helps define confidence intervals (how confident can you be that your results can be extrapolated). One type of random number generator can be found at www.randomizer.org (accessed Mar. 22, 2009). However, there are many different types available online as well as in a variety of reference works.

34. Maki, "Developing an Assessment Plan," 13; indicates that "Initially, limiting the number of outcomes colleagues will assess enables them to determine how an assessment cycle will operate based on existing structures and processes or proposed new ones."

35. Maki, "Developing an Assessment Plan," 11; includes a variety of suggestions for how to establish a schedule for assessment.

36. Matthias Schonlau, Ronald D. Fricker Jr., and Marc N. Elliott, *Conducting Research Surveys via Email and the Web* (Santa Monica, Calif.: Rand, 2002). www.rand.org/publications/MR/MR1480 (accessed Mar. 22, 2009), 40. Schonlau et al. say, "With both volunteer and recruited panels, one concern that researchers have is that participants may tire of filling out surveys, a condition called 'panel fatigue,' or learn to provide the easiest responses, a phenomenon called 'panel conditioning.'"

37. Maki, "Developing an Assessment Plan," 13.

38. Susan J. Beck, "The Extent and Variety of Data-Informed Decisions at Academic Libraries: An Analysis of Nine Libraries of the Association of Research Libraries in the U.S.A. and Canada," *Presentation at the 5th Northumbria International Conference on Performance Measurement in Libraries* (Durham, England: July 29, 2003). Personal paper shared by author. In Beck's presentation she explained that one library "demonstrates its commitment to assessment by spending more than 5 percent of their 17 million dollar ($850,000, 2002 figure) annual budget on [a formal Management Information Systems department staffed with three people]."

39. Hernon and Altman, *Assessing Service Quality,* 124.

40. See also "panel conditioning," Schonlau et al., *Conducting Research Surveys,* 40.

41. Poll and Payne, "Impact Measures," 557–58, describes a similar set of steps for assessment based on an analysis of

a number of institutions who followed a common approach. Hernon and Altman, *Assessing Service Quality*, 47, suggests similar composition to an assessment process, in three parts that are parallel to the question-asking, measurable objectives and standards that I propose.

42. Originally developed for *Assessment Plan for MIT Libraries' Information Services*, http://libstaff.mit.edu/pslg/PS -AssessmentPlanOct2006.pdf (accessed Mar. 22, 2009). The MIT Libraries' original version used the term "benchmarks" rather than "standards," but that term has caused confusion, so I have changed it in this document. These standards were updated in summer 2008, so here I include the updated standards, rather than the originals.

43. For this as well as other information regarding the University of Virginia's balanced scorecard metrics and targets, see www.lib.virginia.edu/bsc/metrics/all0708.html (accessed Mar. 22, 2009), which is specifically information for 2007–2008 at that institution.

44. Hernon and Altman, *Assessing Service Quality*, 67.

45. "Target1 indicates complete success in achieving the metric; Target2 indicates partial success," www.lib.virginia.edu/bsc/metrics/all0708.html (accessed Mar. 22, 2009).

46. Hernon and Altman, *Assessing Service Quality*, 192.

47. Horowitz, *Glossary*.

48 Judy Sharken Simon, "How to Conduct Focus Groups," *Nonprofit World* 17, no. 5 (1999): 40–43. You may wish to read her book, *The Wilder nonprofit field guide to conducting successful focus groups* (Saint Paul, Minn.: Amherst H. Wilder Foundation, 1999).

49. Hernon and Altman, *Assessing Service Quality*, 137ff.

50. Ibid., 100ff.

51. Schonlau et al., *Conducting Research Surveys*.

52 Andrew Hubbertz, "The Design and Interpretation of Unobtrusive Evaluations," *Reference and User Services Quarterly* 44, no. 4 (Summer 2005): 327–35.

53. Hernon and Altman, *Assessing Service Quality*, 195–209.

54. Beck, "Extent and Variety"; Hiller and Self, "From Measurement to Management," 144, also stresses the value of "[supplementing] data with rich qualitative information."

55. Hernon and Altman, *Assessing Service Quality*, 207–08.

56. MIT Libraries, *Assessment Plan*.

Appendix: Performance Guidelines and Standards for Comparison

ALA Standards and Guidelines

Guidelines for Instruction Programs in Academic Libraries, www.ala.org/ala/mgrps/divs/acrl/standards/guidelinesinstruction.cfm (accessed Mar. 22, 2009)

Guidelines for Liaison Work in Managing Collections and Services, www.ala.org/ala/mgrps/divs/rusa/resources/guidelines/guidelinesliaison.cfm (accessed Mar. 22, 2009)

Guidelines for University Library Services to Undergraduate Students, www.ala.org/ala/mgrps/divs/acrl/standards/ulsundergraduate.cfm (accessed Mar. 22, 2009)

Information Literacy Competency Standards for Higher Education, www.ala.org/ala/mgrps/divs/acrl/standards/informationliteracycompetency.cfm (accessed Mar. 22, 2009)

Standards for Libraries in Higher Education, www.ala.org/ala/mgrps/divs/acrl/standards/standardslibraries.cfm (accessed Mar. 22, 2009)

Various guidelines for reference and user services, www.ala.org/ala/mgrps/divs/rusa/resources/guidelines/index.cfm (accessed Mar. 22, 2009; includes behavioral performance guidelines and information competency guidelines)

Other ALA standards and guidelines, www.ala.org/ala/aboutala/offices/ors/standardsa/standardsguidelines/standguide.cfm (accessed Mar. 22, 2009)

Other Standards, Guidelines and Metrics

9 Principles of Good Practice for Assessing Student Learning, AAHE, www.assessment.tcu.edu/assessment/aahe.pdf (accessed Mar. 22, 2009)

Balanced Scorecard at UVa Library, www.lib.virginia.edu/bsc/ (accessed Mar. 22, 2009)