Evaluation Research: An Overview

Ronald R. Powell

Abstract

Evaluation research can be defined as a type of study that uses standard social research methods for evaluative purposes, as a specific research methodology, and as an assessment process that employs special techniques unique to the evaluation of social programs. After the reasons for conducting evaluation research are discussed, the general principles and types are reviewed. Several evaluation methods are then presented, including input measurement, output/performance measurement, impact/outcomes assessment, service quality assessment, process evaluation, benchmarking, standards, quantitative methods, qualitative methods, cost analysis, organizational effectiveness, program evaluation methods, and LIS-centered methods. Other aspects of evaluation research considered are the steps of planning and conducting an evaluation study and the measurement process, including the gathering of statistics and the use of data collection techniques. The process of data analysis and the evaluation report are also given attention. It is concluded that evaluation research should be a rigorous, systematic process that involves collecting data about organizations, processes, programs, services, and/or resources. Evaluation research should enhance knowledge and decision making and lead to practical applications.

What Is Evaluation Research?

Evaluation research is not easily defined. There is not even unanimity regarding its name; it is referred to as *evaluation* research and *evaluative* research. Some individuals consider evaluation research to be a specific research method; others focus on special techniques unique, more often

LIBRARY TRENDS, Vol. 55, No. 1, Summer 2006 ("Research Methods," edited by Lynda M. Baker), pp. 102–120

than not, to program evaluation; and yet others view it as a research activity that employs standard research methods for evaluative purposes. Consistent with the last perspective, Childers concludes, "The differences between evaluative research and other research center on the orientation of the research and not on the methods employed" (1989, p. 251). When evaluation research is treated as a research method, it is likely to be seen as a type of applied or action research, not as basic or theoretical research.

Weiss, in her standard textbook, defines evaluation as "the *systematic assessment* of the *operation* and/or the *outcomes* of a program or policy, compared to a set of *explicit* or *implicit standards,* as a means of contributing to the *improvement* of the program or policy" (1998, p. 4; emphasis in original). While certainly not incorrect, this definition, at least within a library and information (LIS) context, is too narrow or limited. Wallace and Van Fleet, for example, point out that "evaluation has to do with understanding library systems" (2001, p. 1). As will be noted later in this article, evaluative methods are used for everything from evaluating library collections to reference transactions.

Why Evaluate?

But before examining the specific techniques and methods used in LIS evaluation research, let us first briefly consider the question of why evaluation is important and then identify the desirable characteristics of evaluation, the steps involved in planning an evaluation study, and the general approaches to evaluation. With regard to the initial question, Wallace and Van Fleet (2001, pp. xx-xxi) and others have noted that there are a growing number of reasons why it is important for librarians and other information professionals to evaluate their organizations' operations, resources, and services. Among those reasons are the need for organizations to

1. account for how they use their limited resources
2. explain what they do
3. enhance their visibility
4. describe their impact
5. increase efficiency
6. avoid errors
7. support planning activities
8. express concern for their public
9. support decision making
10. strengthen their political position.

In addition to some of the reasons listed above, Weiss (1998, pp. 20–28) identifies several other purposes for evaluating programs and policies. They include the following:

1. Determining how clients are faring
2. Providing legitimacy for decisions
3. Fulfilling grant requirements
4. Making midcourse corrections in programs
5. Making decisions to continue or culminate programs
6. Testing new ideas
7. Choosing the best alternatives
8. Recording program history
9. Providing feedback to staff
10. Highlighting goals

"Over the past decade, both academics and practitioners in the field of library and information science (LIS) have increasingly recognized the significance of assessing library services" (Shi & Levy, 2005, p. 266). In August 2004 the National Commission on Libraries and Information Science announced three strategic goals to guide its work in the immediate future. Among those three goals was the appraising and assessing of library and information services.

Characteristics and Principles of Evaluation

Childers (1989, p. 250), in an article emphasizing the evaluation of programs, notes that evaluation research (1) is usually employed for decision making; (2) deals with research questions about a program; (3) takes place in the real world of the program; and (4) usually represents a compromise between pure and applied research. Wallace and Van Fleet (2001) comment that evaluation should be carefully planned, not occur by accident; have a purpose that is usually goal oriented; focus on determining the quality of a product or service; go beyond measurement; not be any larger than necessary; and reflect the situation in which it will occur. Similarly, evaluation should contribute to an organization's planning efforts; be built into existing programs; provide useful, systematically collected data; employ an outside evaluator/consultant when possible; involve the staff; not be any fancier than necessary; and target multiple audiences and purposes (Some Practical Lessons on Evaluation, 2000).

In an earlier work on the evaluation of special libraries, Griffiths and King (1991, p. 3) identify some principles for good evaluation that still bear repeating:

1. Evaluation must have a purpose; it must not be an end in itself
2. Without the potential for some action, there is no need to evaluate
3. Evaluation must be more than descriptive; it must take into account relationships among operational performance, users, and organizations
4. Evaluation should be a communication tool involving staff and users
5. Evaluation should not be sporadic but be ongoing and provide a means for continual monitoring, diagnosis, and change

6. Ongoing evaluation should provide a means for continual monitoring, diagnosis and change
7. Ongoing evaluation should be dynamic in nature, reflecting new knowledge and changes in the environment

As has been implied, but not explicitly stated above, evaluation often attempts to assess the *effectiveness* of a program or service. On a more specific level, evaluation can be used to support accreditation reviews, needs assessments, new projects, personnel reviews, conflict resolution, and professional compliance reports.

Types of Evaluation Research

Before selecting specific methods and data collection techniques to be used in an evaluation study, the evaluator, according to Wallace and Van Fleet (2001), should decide on the general approach to be taken. They categorize the general approaches as ad hoc/as needed/as required or evaluation conducted when a problem arises; externally centered, or evaluation necessitated by the need to respond to external forces such as state library and accrediting agencies; internally centered, or evaluation undertaken to resolve internal problems; and research centered, or evaluation that is conducted so that the results can be generalized to similar environments. Other broad categories of evaluation that can encompass a variety of methods include macroevaluation, microevaluation, subjective evaluation, objective evaluation, formative evaluation (evaluation of a program made while it is still in progress), and summative evaluation (performed at the end of a program). The *Encyclopedia of Evaluation* (Mathison, 2004) treats forty-two different evaluation approaches and models ranging from "appreciative inquiry" to "connoisseurship" to "transformative evaluation."

Evaluation Methods

Having decided on the general approach to be taken, the evaluator must next select a more specific approach or method to be used in the evaluation study. What follows are brief overviews of several commonly used evaluation methods or groups of methods.

Input Measurement

Input measures are measures of the resources that are allocated to or held by an organization and represent the longest-standing, most traditional approach to assessing the quality of organizations and their resources and services. Examples of input measures for libraries include the number of volumes held, money in the budget, and number of staff members. By themselves they are more measurement than true evaluation and are limited in their ability to assess quality.

Output/Performance Measurement

Output or performance measures serve to indicate what was accomplished as a result of some programmatic activity and thus warrant being considered as a type of evaluation research. Such measures focus on indicators of library output and effectiveness rather than merely on input; are closely related to the impact of the library on its community; and, as is true for virtually all evaluation methods, should be related to the organization's goals and objectives.

As was just indicated, one critical element of performance measurement is effectiveness; another is user satisfaction. In addition to user satisfaction, examples of performance/output measures include use of facilities and equipment, circulation of materials, document delivery time, reference service use, subject search success, and availability of materials. The Association of Research Libraries (2004) identified the following eight output measures for academic libraries: ease and breadth of access, user satisfaction, teaching and learning, impact on research, cost effectiveness of library operations and services, facilities and space, market penetration, and organizational capacity. One could argue that not all of those eight measures represent true performance or output measures, but they are definitely measures of effectiveness.

Impact/Outcomes Assessment

The input or resources of a library are relatively straightforward and easy to measure. True measurement of the performance of a library is more difficult to achieve, and it is even more challenging to measure impact/outcomes or how the use of library and information resources and services actually affects users. Rossi, Lipsey, and Freeman (2004) point out that outcomes must relate to the benefits of products and services, not simply their receipt (a performance measure). However, given the increasing call for accountability, it is becoming imperative for libraries to measure outcomes or impact. Indeed, "outcomes evaluation has become a central focus, if not the central focus, of accountability-driven evaluation" (Patton, 2002, p. 151).

Some authors use the terms *impact* and *outcome* synonymously; others see them as somewhat different concepts. Patton (2002, p. 162) suggests a logical continuum that includes inputs, activities and processes, outputs, immediate outcomes, and long-term impacts. Bertot and McClure, in a 2003 article in *Library Trends* (pp. 599–600), identified six types of outcomes:

1. Economic: outcomes that relate to the financial status of library users
2. Learning: outcomes reflecting the learning skills and acquisition of knowledge of users
3. Research: outcomes that include, for example, the impacts of library services and resources on the research process of faculty and students

4. Information Exchange: outcomes that include the ability of users to exchange information with organizations and other individuals
5. Cultural: the impact of library resources and services on the ability of library users to benefit from cultural activities
6. Community: outcomes that affect a local community and in turn affect the quality of life for members of the community

Matthews (2004, pp. 109–110), in his book on measuring public library effectiveness, identifies six categories of outcomes or benefits for public libraries. Those six categories, with examples, are as follows:

1. Cognitive results: refreshed memory, new knowledge, changed ideas
2. Affective results: sense of accomplishment, sense of confidence
3. Meeting expectations: getting what they needed, getting too much, seeking substitute sources
4. Accomplishments: able to make better-informed decisions, achieving a higher quality performance
5. Time aspects: saved time, wasted time, had to wait for service
6. Money aspects: the dollar value of results obtained, the amount of money saved, the cost of using the service

Impacts more relevant to academic libraries and their users include improved test scores, better papers, publications, increased class participation, etc. (Powell, 1995). A book by Hernon and Dugan (2002) considers outcomes for both academic and public libraries. The latter include getting ideas, making contact with others, resting or relaxing, and being entertained. Markless and Streatfield (2001) examine impact indicators for public, school, and academic libraries. Among their impact targets for school libraries are "improved quality and type of communication between learners and LRC staff" and "enhanced user confidence" (p. 175). Seadle (2003) notes that outcome-based evaluation is increasingly used for digital library projects.

Service Quality

Service quality, briefly defined, is "the difference between a library user's expectations and perceptions of service performance" (Nitecki, 1996, p. 182). As a concept, it dates back to at least the 1970s and has some roots in the total quality management (TQM) movement. TQM is characterized by the implementation of standards of quality, the encouragement of innovation, the measurement of results, and the taking of corrective actions as needed. TQM emphasizes the use of a team approach to maximizing customer satisfaction. A 1996 article by Pritchard provides an excellent overview of TQM, as well as other approaches to determining quality.

Quality is an elusive concept for which there is no commonly accepted definition, but the assessment of service quality did get a boost from earlier research from Parasuraman, Berry, and Zeithaml (see Nitecki, 1996). They

developed a conceptual framework, the Gaps Model of Service Quality, and a widely used instrument, SERV-QUAL, for measuring service quality. The Gaps Model incorporates the following gaps, as measured by the SERV-QUAL questionnaire:

1. The discrepancy between customers' expectations and managements' perceptions of these expectations
2. The discrepancy between managements' perceptions of customers' expectations and service-quality specifications
3. The discrepancy between service-quality specifications and actual service delivery
4. The discrepancy between actual service delivery and what is communicated to customers about it
5. The discrepancy between customers' expected services and perceived services delivered (Nitecki, 1996, p. 182)

The most visible current iteration of SERV-QUAL in the library field is known as LibQUAL+. LibQUAL+ was developed by faculty members of Texas A&M University in partnership with the Association of Research Libraries (ARL) and is part of ARL's New Measures Initiative. Over the past few years LibQUAL+ studies have been conducted by hundreds of libraries, including many large university libraries in the United States. These studies are intended for libraries "to solicit, track, understand, and act upon users' opinions of service quality" (*LibQUAL+*, 2003). Questions in the LibQUAL+ questionnaire address library staff, print and electronic resources, service hours, facilities, equipment, and document delivery and gather the data needed to calculate the gaps described above. However, according to Shi and Levy, "the current LibQUAL+ is not yet an adequately developed tool to measure and represent a dependable library services assessment result" (2005, p. 272).

Individuals wanting to know more about the use of service quality methods in academic libraries may wish to read a book by Hernon and Altman (1996). Other models of quality assessment from a British perspective are considered by Jones, Kinnell, and Usherwood (2000).

Process Evaluation

The second stage in Patton's (2002) continuum described in the section on impact/outcomes assessment was *processes* or activities. "A focus on process involves looking at *how* something happens rather than or in addition to examining outputs and outcomes" (p. 159). "Process data permit judgments about the extent to which the program or organization is operating the way it is supposed to be operating, revealing areas in which relationships can be improved as well as highlighting strengths of the program that should be preserved" (Patton, 2002, p. 160). Process evaluation focuses on "what the

program actually *does*" (Weiss, 1998, p. 9). It "is the most frequent form of program evaluation" (Rossi, Lipsey, & Freeman, 2004, p. 57).

Process indicators are somewhat similar to performance measures, but they focus more on the activities and procedures of the organization than on the products of those activities. For example, a process evaluation of an acquisitions department would be concerned with how materials are acquired and prepared for the shelf, not on how many books are ultimately used. In an academic library setting, process indicators might include staff training and development, delivery styles, knowledge of the curriculum, and participation in assignments and grading (Markless & Streatfield, 2001). In his book on public library effectiveness, Matthews (2004) places process measures in three categories: efficiency, staff productivity, and library information system activity. More generally speaking, a process evaluation "might examine how consistent the services actually delivered are with the goals of the program, whether services are delivered to appropriate recipients, how well service delivery is organized, the effectiveness of program management, the use of program resources, and other such matters" (Rossi, Lipsey, & Freeman, 2004, p. 57). And ultimately, the evaluator would want to know the extent to which programs and services were actually implemented. Patton (2002) even argues that "implementation evaluation" is a distinct method, and in many cases implementation information is of greater value than outcomes information (p. 161).

Benchmarking

One of the relatively recent approaches to measuring the performance of libraries and other organizations is benchmarking. Benchmarking tends to fall into the "total quality management" category. Benchmarking "represents a structured, proactive change effort designed to help achieve high performance through comparative assessment. It is a process that establishes an external standard to which internal operations can be compared" (Jurow, 1993, p. 120). The 2000 *Standards for College Libraries* describes benchmarking as the process of evaluating a library's points of comparison—inputs and outputs—against its peers and aspirational peers. There are several types of benchmarking, one of which is referred to as competitive or performance benchmarking. Performance benchmarking utilizes comparative data gathered from the same field or the same type of organization. The data are usually derived from analyses of organizational processes and procedures. Benchmarking can be used to establish best practices, identify changes to improve services, evaluate opinions and needs of users, identify trends, exchange ideas, and develop staff. Peischl (1995) points out that candidates for benchmarking include the services or products of an organization, internal work processes, internal support functions, and organizational performance and strategy.

Standards

According to Baker and Lancaster, "standards have an important role to play in the evaluation of library services . . . When applied to libraries, however, *standards* refers to a set of guidelines or recommended practices, developed by a group of experts, that serve as a model for good library service" (1991, p. 321). Some general types of standards, as identified by Baker and Lancaster (1991), include technical standards (for example, cataloging codes), performance standards, output measures, input measures, qualitative standards, and quantitative standards.

Quantitative Evaluation

Any evaluation method that involves the measurement of quantitative/numerical variables probably qualifies as a quantitative method, and many of the methods already examined fall into this broad category. Among the strengths of quantitative methods are the evaluator can reach conclusions with a known degree of confidence about the extent and distribution of that the phenomenon; they are amenable to an array of statistical techniques; and they are generally assumed to yield relatively objective data (Weiss, 1998, pp. 83–84).

Experimental methods usually, but not always, deal with quantitative data and are considered to be the best method for certain kinds of evaluation studies. Indeed, "the classic design for evaluations has been the experiment. It is the design of choice in many circumstances because it guards against the threats to validity" (Weiss, 1998, p. 215). The experiment is especially useful when it is desirable to rule out rival explanations for outcomes. In other words, if a true experimental design is used properly, the evaluator should be able to assume that any net effects of a program are due to the program and not to other external factors.

On the other hand, experimental methods are relatively weak in producing findings that can be generalized to other situations because they are usually conducted in rather controlled settings. Also, experiments tend to be used to test the effects of one component of a program at a time rather than the entire program. Another limitation of the true or randomized experiment is that it is not well suited for evaluating programs in their early stages of implementation. If the program changes significantly before outcomes are measured, it will be difficult to determine which version of the program produced what effects (Rossi, Lipsey, & Freeman, 2004).

Survey methods are often quantitative in nature but lack the experiment's ability to rigorously test the relationship between a program or service and its outputs or impact. Questionnaires and interviews, and observation to a lesser degree, represent the most commonly used survey data gathering techniques. Other quantitative methods covered by the *Encyclopedia of Evaluation* (Mathison, 2004) include concept mapping, correlation, cross-sectional design, matrix sampling, meta-analysis, panel studies, regression analysis, standardized tests, and time series analysis.

Qualitative Evaluation

As is true for basic research, qualitative methods are becoming increasingly popular. In fact, "the most striking development in evaluation in recent years is the coming of age of qualitative methods. Where once they were viewed as aberrant and probably the refuge of those who had never studied statistics, now they are recognized as valuable additions to the evaluation repertoire" (Weiss, 1998, p. 252). The *Encyclopedia of Evaluation* (Mathison, 2004) includes thirty-seven qualitative methods. They are appropriate, of course, when the phenomena being evaluated do not lend themselves to quantification. A qualitative method "tends to apply a more holistic and natural approach to the resolution of the problem than does quantitative research. It also tends to give more attention to the subjective aspects of human experience and behavior" (Powell & Connaway, 2004, p. 59). "Qualitative strategies can be particularly appropriate where the administration of standardized instruments, assigning people to comparison groups [in experiments], and/or the collection of quantitative data would affect program operations by being overly intrusive" (Patton, 2002, p. 191). In addition, they can provide

1. greater awareness of the perspective of program participants and often a greater responsiveness to their interests
2. capability for understanding dynamic developments in the program as it evolves
3. awareness of time and history
4. special sensitivity to the influence of context
5. ability to enter the program scene without preconceptions or prepared instruments, and to learn what is happening
6. alertness to unanticipated and unplanned events
7. general flexibility of perspective (Weiss, 1998, p. 253).

Qualitative methods do have their disadvantages as well, of course. Among them are the following:

1. Limited ability to yield objective data
2. Limited ability to produce generalizable results
3. Limited ability to provide precise descriptions of program outcomes
4. Not well suited for developing *specific* answers about the relationship of particular program strategies or events to outcomes (Weiss, 1998, pp. 85–86)
5. Often relatively labor intensive to conduct

Cost Analysis

Simple cost analysis is basically a descriptive breakdown of the costs incurred in operating an organization. Cost-related techniques more concerned with the assessment of whether monies are being spent in an optimal fashion usually fall into one of two groups—cost-effectiveness studies and

cost-benefit analysis. "The term 'cost-effectiveness' implies a relationship between the cost of providing some service and the level of effectiveness of that service . . . Cost-effectiveness analyses can be thought of as studies of the costs associated with alternative strategies for achieving a particular level of effectiveness" (Lancaster, 1993, p. 267). Some examples of cost-effectiveness measures include the cost per relevant informational resource retrieved, cost per use of a resource, cost per user, cost per capita, and cost by satisfaction level (Lancaster, 1993; Matthews, 2004).

Cost-effectiveness analysis can be seen as "a truncated form of cost-benefit analysis that stops short of putting an economic value on . . . outcomes [benefits] of programs" (Klarman, 1982, p. 586). "'Cost-benefit,' clearly, refers to a relationship between the cost of some activity and the benefits derived from it. In effect, a cost-benefit study is one that tries to justify the existence of the activity by demonstrating that the benefits outweigh the costs" (Lancaster, 1993, p. 294). A typical cost-benefit analysis involves determining who benefits from and pays for a service, identifying the costs for each group of beneficiaries, identifying the benefits for each group, and comparing costs and benefits for each group to determine if groups have net benefits or net costs and whether the total benefits exceed the total costs.

Types of cost-benefit analysis described by Lancaster (1993) are

1. net value approach: the maximum amount the user of an information service is willing to pay minus the actual cost
2. value of reducing uncertainty in decision making
3. cost of buying service elsewhere
4. librarian time replaces user time (that is, the librarian saves the user time by performing his or her task)
5. service improves organization's performance or saves it money.

Other kinds of cost analysis discussed by Weiss (1998) and Matthews (2004) include the following:

1. Cost-minimization analysis: seeks to determine the least expensive way to accomplish some outcome
2. Cost-utility analysis: considers the value or worth of a specific outcome for an individual or society
3. Willingness-to-pay approach: asks how much individuals are willing to pay to have something they currently do not have
4. Willingness-to-accept approach: asks individuals how much they would be willing to accept to give up something they already have
5. Cost of time

Organizational Effectiveness

The determination of the effectiveness of an organization has been identified as one of the objectives for some of the methods described above, and, indeed, it may be more properly thought of as an evaluation objective than an evaluation method. Regardless, it is a crucial element of organizational assessment and has received considerable attention in the professional literature. Rubin (cited by Wallace and Van Fleet, 2001, pp. 13–14) identifies a number of criteria for effectiveness at the organizational level and then describes several models for measuring organizational effectiveness. Those models and their "key questions" are as follows:

1. Goals: Have the established goals of the library been met?
2. Critical Constituencies: Have the needs of constituents been met?
3. Resources: Have necessary resources been acquired?
4. Human Resources: Is the library able to attract, select, and retain quality employees?
5. Open Systems: Is the library able to maintain the system, adapt to threats, and survive?
6. Decision Process: How are decisions made and evaluated?
7. Customer Service: How satisfied is the clientele with the library?

Program Evaluation Methods

In addition to the methods already identified, there are numerous other methods primarily used for social program evaluation. Readers interested in learning more about such methods are referred to the works on evaluation already cited above, including the article by Childers (1989), and to the table by King in Powell and Connaway (2004, pp. 57–58).

LIS-Centered Methods

Another approach to categorizing evaluation methods used in library and information science is according to the program, service, or resource to be evaluated. The book by Wallace and Van Fleet (2001), for example, has chapters devoted to the evaluation of reference and information services and to library collections (see Whitlatch, 2001 for an article on the evaluation of electronic reference services). Bawden (1990) presents a user-oriented approach for the evaluation of information systems and services. An earlier issue of *Library Trends* (Reed, 1974) has articles on the evaluation of administrative services, collections, processing services, adult reference service, public services for adults, public library services for children, and school library media services. Lancaster's 1993 text includes the evaluation of collections, collection use, in-house library use, periodicals, library space, catalog use, document delivery, reference services, and resource sharing. Most of these methods, however, actually employ techniques related to the more generic methods identified earlier in this article.

Planning the Evaluation Study

As has already been indicated, evaluation should be part of an organization's overall planning process and integral to the assessment of current services and resources, the development of strategies for change, and the monitoring of progress toward goals and objectives. Indeed, in order to be valid, an evaluation must reflect the organization's mission, goals, and objectives. In planning the evaluation of a specific program, the evaluator should first gather relevant background information. This activity might well include reviewing the professional literature, identifying professional standards and guidelines, and networking with colleagues. Next, the evaluator should decide what he or she actually wants to know, that is, focus the evaluation. This requires a determination of the purpose(s) of the evaluation specific to the program being examined. For example, the purpose may simply be to learn more about the program, or it may be to determine if the program is meeting its objectives.

After focusing the evaluation, decisions must be made about the overall design of the study, the method(s) to be used, and the measurements to be made. In other words, the evaluator must decide what must be measured, choose an evaluation method, select the data collection techniques to be employed, plan the construction and/or purchase of data collection instruments, plan the data analysis, develop a budget for the evaluation study, and recruit personnel. As is often the case in research studies, it is a good idea to utilize more than one method so as to increase the reliability and validity of the study and its findings. Haynes (2004, p. 19), for example, argues for mixed-method evaluation, which combines user-centered with system-centered paradigms and qualitative with quantitative methods. It is a good idea to write a thorough plan or proposal for the study at this time.

Weiss (1998) reminds us that the evaluator should also give careful thought to the best time to conduct the evaluation, the types of questions to ask, whether one or a series of studies will be necessary, and any ethical issues that might be generated by the study. Those and other planning points are succinctly represented in the following "evaluation action plan" suggested by Wallace and Van Fleet (2001, pp. 4–5):

1. What's the problem?
2. Why am I doing this?
3. What *exactly* do I want to know?
4. Does the answer already exist?
5. How do I find out?
6. Who's involved?
7. What's this going to cost?
8. What will I do with the data?
9. Where do I go from here?

Conducting the Evaluation Study

After planning the evaluation, it is time, of course, to conduct the study. That is, the evaluator is now ready to collect data or measure what needs to be measured; analyze the data; and report the findings. What follows is a brief overview of the steps in the evaluation process.

Measurement

"Measurement, in most general terms, can be regarded as the assignment of numbers to objects (or events or situations) in accord with some rule. The property of the objects which determines the assignment according to that rule is called *magnitude,* the measurable attribute; the number assigned to a particular object is called its *measure,* the amount or degree of its magnitude" (Kaplan, 1964, p. 177). More generally, measurement is any process for describing in quantitative values things, people, events, etc. Measurement by itself is not true evaluation, but it is one of the building blocks for quantitative evaluation. Common types of measures for library evaluation studies include number and types of users, number and duration of transactions, user and staff activities, user satisfaction levels, and costs of resources and services. They can be related to input, output, effectiveness, costs, etc.

It is critical that the measurement process and the measures be reasonably high in reliability and validity. Reliability refers to the degree to which measurements can be depended upon to secure consistent and accurate results in repeated applications. Validity is the degree to which any measure or data collection technique succeeds in doing what it purports to do; it refers to the meaning of an evaluative measure or procedure. The validity and/or reliability of measures can be affected by such factors as inconsistent data collection techniques, biases of the observer, the data collection setting, instrumentation, behavior of human subjects, and sampling. The use of multiple measures can help to increase the validity and reliability of the data. They are also worth using because no single technique is up to measuring a complex concept, multiple measures tend to complement one another, and separate measures can be combined to create one or more composite measures (Weiss, 1998).

Statistics

Many measures are in the form of statistics, which, in some cases, can be drawn from already existing sources of data. Types of statistics include administrative data, financial statistics, collections and other resources or inputs, use and other output/performance measures, outcomes, and staff and salary information. Sources of statistics include governmental agencies, professional associations, and other organizations such as state library agencies. Among the noteworthy sources of library-related statistics are the National Center for Education Statistics (NCES), American Library Association and its divisions (such as the Public Library Association's Public

Library Data Service and the Association of College and Research Libraries' Trends and Statistics series), Association of Research Libraries, and federal programs such as the Federal State Cooperative System and the Integrated Postsecondary Education Data System.

Data Collection Techniques

The evaluator must next select or design one or more data collection techniques that are compatible with the method(s) to be used and that are capable of gathering the necessary information. There are too many data collection techniques to consider here, but some of the relatively common techniques and instruments used for evaluation studies, as well as for other kinds of research, include the following:

1. Tests (standardized and locally developed)
2. Assessments by participants
3. Assessments by experts
4. Questionnaires (paper and electronic)
5. Interviews, including focus groups
6. Observation of behavior and activities
7. Evaluation of staff performance
8. Analysis of logs or diaries of participants
9. Analysis of historical and current records
10. Transactional log analysis
11. Content analysis
12. Bibliometrics, especially citation analysis
13. Use records
14. Anecdotal evidence

For information about many of these techniques, readers are referred to Powell and Connaway (2004) and Hernon and McClure (1990). For more information about techniques unique to evaluations of library and information use, readers may wish to consult earlier texts by Lancaster (1993) and Baker and Lancaster (1991). Westbrook's chapter in Powell and Connaway (2004), a chapter in Weiss (1998), and the book by Patton (2002) are among the sources of information about qualitative data collection techniques.

Analysis of Data

"The aim of analysis is to convert a mass of raw data into a coherent account. Whether the data are quantitative or qualitative, the task is to sort, arrange, and process them and make sense of their configuration. The intent is to produce a reading that accurately represents the raw data and blends them into a meaningful account of events" (Weiss, 1998, p. 271). The basic tasks of data analysis for an evaluative study are to answer the questions that must be answered in order to determine the success of the program or service, the quality of the resources, etc. Those questions

should, of course, be closely related to the nature of what is being evaluated and the goals and objectives of the program or service. In addition, the nature of the data analysis will be significantly affected by the methods and techniques used to conduct the evaluation. According to Weiss (1998), most data analyses, whether quantitative or qualitative in nature, will employ some of the following strategies: describing, counting, factoring (that is, dividing into constituent parts), clustering, comparing, finding commonalities, examining deviant cases, finding covariation, ruling out rival explanations, modeling, and telling the story. Evaluators conducting quantitative data analyses will need to be familiar with techniques for summarizing and describing the data (that is, descriptive statistics); and if they are engaged in testing relationships or hypotheses and/or generalizing findings to other situations, they will need to utilize inferential statistics.

Whatever the nature of the data analysis, however, it cannot substitute for sound development of the study and interpretation of the findings. Statistics can only facilitate the interpretation. In a quantitative study the analysis and interpretation usually follow the conduct of the study. In a qualitative study the data analysis is typically concurrent with the data gathering; "nor, in practice, are analysis and interpretation neatly separated" (Patton, 1987, p. 144).

The Evaluation Report

As part of the planning, the evaluator should have considered how and to whom the findings will be communicated and how the results will be applied. Weiss (1998, pp. 296–297) recommends that the typical report of a program evaluation include the following elements:

1. Summary of study results
2. Problem with which the program deals
3. Nature of the program: goals and objectives, activities, context, beneficiaries, staff
4. Nature of the evaluation
5. Comparison with evaluations of similar programs (optional)
6. Suggestions for further evaluation (optional)

A good report will be characterized by clarity, effective format and graphics, timeliness, candor about strengths and weaknesses of the study, and generalizability (Weiss, 1998), as well as by adequacy of sources and documentation, appropriateness of data analysis and interpretation, and basis for conclusions.

Conclusions

As was indicated above, evaluation research has been defined in a number of ways. It is viewed as a specific research methodology, as a type of study that uses standard social research methods for evaluative purposes,

and as an assessment process employing special techniques unique to the evaluation of programs. If treated as research, it is likely to be designed as applied or action research even though it may well use basic research methods. But generally speaking, all of the approaches to evaluation tend to share the following important commonalities: evaluation is a systematic process; it involves collecting data about organizations, processes, programs, services, and resources; it is a process for enhancing knowledge and decision making; and it is expected to lead to practical applications (Preskill & Russ-Eft, 2005, pp. 1–2). And finally, evaluation research should be conducted carefully and rigorously with consideration of many of the tenets that characterize good basic research.

References

Association of Research Libraries. (2004). *ARL new measures retreat; Transcript.* Washington, DC: Association of Research Libraries.

Baker, S. L., & Lancaster, F. W. (1991). *The measurement and evaluation of library services* (2nd ed.). Arlington, VA: Information Resources Press.

Bawden, D. (1990). *User-oriented evaluation of information systems and services.* Brookfield, VT: Gower.

Bertot, J. C., & McClure, C. R. (2003). Outcomes assessment in the networked environment: Research questions, issues, considerations, and moving forward. *Library Trends, 51*(4), 590–613.

Childers, T. (1989). Evaluative research in the library and information field. *Library Trends, 38*(2), 250–267.

Griffiths, J. M., & King, D. W. (1991). *A manual on the evaluation of information centers and services.* New York: American Institute of Aeronautics and Astronautics Technical Information Service.

Haynes, A. (2004). Bridging the gulf: Mixed methods and library service evaluation. *Australian Library Journal, 53*(3), 285–307.

Hernon, P., & Altman, E. (1996). *Service quality in academic libraries.* Norwood, NJ: Ablex.

Hernon, P., & Dugan, R. E. (2002). *An action plan for outcomes assessment in your library.* Chicago: American Library Association.

Hernon, P., & McClure, C. R. (1990). *Evaluation and library decision making.* Norwood, NJ: Ablex.

Jones, K., Kinnell, M., & Usherwood, B. (2000). The development of self-assessment tool-kits for the library and information sector. *Journal of Documentation, 56*(2), 119–135.

Jurow, S. R. (1993). Tools for measuring and improving performance. In S. Jurow & S. B. Barnard (Eds.), *Integrating Total Quality Management in a library setting* (pp. 113–126). New York: Haworth.

Kaplan, A. (1964). *The conduct of inquiry: Methodology for behavioral science.* San Francisco: Chandler.

Klarman, H. E. (1982). The road to cost-effectiveness analysis. *Milbank Memorial Fund Quarterly: Health and Society, 60*(4), 585–603.

Lancaster, F. W. (1993). *If you want to evaluate your library . . .* (2nd ed.). Champaign, IL: University of Illinois, Graduate School of Library and Information Science.

LibQUAL+ Spring 2003 survey results. (2003). Washington, DC: Association of Research Libraries.

Markless, S., & Streatfield, D. (2001). Developing performance and impact indicators and targets in public and education libraries. *International Journal of Information Management, 21*(2), 167–179.

Mathison, S. (Ed.). (2004). *Encyclopedia of evaluation.* Thousand Oaks, CA: Sage Publications.

Matthews, J. R. (2004). *Measuring for results: The dimensions of public library effectiveness.* Westport, CN: Libraries Unlimited.

Nitecki, D. A. (1996). Changing the concept and measure of service quality in academic libraries. *Journal of Academic Librarianship, 22*(3), 181–190.
Patton, M. Q. (1987). *How to use qualitative methods in evaluation.* Newbury Park, CA: Sage Publications.
Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.
Peischl, T. M. (1995). Benchmarking: A process for improvement. *Proceedings of the First International Conference on TQM and Academic Libraries* (pp. 119–122). Washington, DC: Association of Research Libraries
Powell, R. R. (1995). Impact assessment of university libraries. In *Encyclopedia of library and information science* (Vol. 55, pp. 151–164). New York: Marcel Dekker.
Powell, R. R., & Connaway, L. S. (2004). *Basic research methods for librarians* (4th ed.). Westport, CN: Libraries Unlimited.
Preskill, H., & Russ-Eft, D. (2005). *Building evaluation capacity: 72 activities for teaching and training.* Thousand Oaks, CA: Sage Publications.
Pritchard, S. M. (1996). Determining quality in academic libraries. *Library Trends, 44*(3), 572–594.
Reed, S. R. (Ed.). (1974). Evaluation of library service [Special Issues]. *Library Trends, 22*(3).
Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage Publications.
Seadle, M. (2003). Editorial: Outcome-based evaluation. *Library Hi Tech, 21*(1), 5–7.
Shi, X., & Levy, S. (2005). A theory-guided approach to library services assessment. *College & Research Libraries, 66*(3), 266–277.
Some practical lessons on evaluation. (2000). *Snapshots: Research Highlights from the Nonprofit Sector Research Fund,* No. 9.
Standards for College Libraries. (2000). Chicago: American Library Association.
Wallace, D. P., & Van Fleet, C. (2001). *Library evaluation: A casebook and can-do guide.* Englewood, CO: Libraries Unlimited.
Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
Whitlatch, J. B. (2001). Evaluating reference services in the electronic age. *Library Trends, 50*(2), 207–217.

Additional References

Abels, E. G., Kantor, P. B., & Saracevic, T. (1996). Studying the cost and value of library and information services. *Journal of the American Society for Information Science, 47*(3), 217–227.
Bertot, J. C., McClure, C. R., & Ryan, J. (2001). *Statistics and performance measures for public library networked services.* Chicago: American Library Association.
Childers, T. A., & Van House, N. A. (1993). *What's good? Describing your public library's effectiveness.* Chicago: American Library Association.
Everhart, N. (1998). *Evaluating the school library media center: Analysis techniques and research practices.* Englewood, CO: Libraries Unlimited.
Nelson, W. N., & Fernekes, R. W. (2002). *Standards and assessment for academic libraries: A workbook.* Chicago: Association of College and Research Libraries.
Nisonger, T. E. (2003). *Evaluation of library collections, access and electronic resources: A literature guide and annotated bibliography.* Westport, CN: Libraries Unlimited.
Smith, M. L. (1996). *Collecting and using public library statistics: A how-to-do-it manual for librarians.* New York: Neal-Schuman.
Van House, N. A., Lynch, M. J., McClure, C. R., Zweizig, D. L., & Rodger, E. J. (1987). *Output measures for public libraries* (2nd ed.). Chicago: American Library Association.
Van House, N., Weil, B, & McClure, C. (1990). *Measuring academic library performance: A practical approach.* Chicago: American Library Association.
Whitlatch, J. B. (2000). *Evaluating reference services: A practical guide.* Chicago: American Library Association.
Zweizig, D. (Ed.). (1994). *The tell it! manual: The complete program for evaluating library performance.* Chicago: American Library Association.

Ronald R. Powell is a professor in the Library and Information Science Program at Wayne State University, Detroit, Michigan. Prior to becoming a library and information science educator, he served as a university librarian and college library director. Dr. Powell has taught, conducted research, and published in the areas of research methods, collection development, bibliographic instruction, academic libraries, the measurement and evaluation of library resources and services, and education for librarianship. His publications include *Basic Research Methods for Librarians* (with Lynn Connaway), *Basic Reference Sources* (with Margaret Taylor), *Qualitative Research in Information Management* (with Jack Glazier), and *The Next Library Leadership* (with Peter Hernon and Arthur Young).