Assessing Service to Special Populations

ABSTRACT

Over the past twenty years, librarianship has promulgated quantitative evaluation through the application of output measures to a goal-based model, even in the face of evidence that such an approach makes difficult the fair assessment of services to special populations. While outside librarianship the emphasis is on outcome measurement, we have failed to move into that realm, even when it is most appropriate. In the future, the way in which evaluation is conducted must be determined by the questions it seeks to answer, the model that will best supply the answers, and the design that will uncover an accurate reflection of the program. That requires a combination of qualitative and quantitative measurement rigorously applied. Eight models are suggested that can provide the valid, reliable evaluations that have to date eluded us.

INTRODUCTION

Not unlike other professions, librarianship has resisted evaluation. At the federal level, even with legislation like the Library Services and Construction Act (LSCA) Title I, which has as its major focus service to special populations—the aging, handicapped, disadvantaged minorities, the illiterate, and those for whom English is a second language—hard-hitting comments have become part of the record on library efforts (Shavitt, 1985, pp. 124-25). Although assessment is required to receive LSCA funding, the consensus of recent studies,
including a 1989 meta-evaluation, is that library program evaluation stands now where educational program evaluation stood fifteen years ago (Roberts, 1985, p. 1; Turock, 1990, p. 50).

**Why Is Evaluation Resisted?**

Given this negativity, why do librarians continue to resist evaluation? Frequently, that question is answered by citing a tradition of limited interest which, in turn, is blamed on a limited knowledge and understanding of evaluation processes and techniques. But that supposition is not only condescending, it also reinforces the unrealistic expectation that minimal knowledge of the evaluative process will not harm the validity of the resulting product.

At a Midwinter Conference held during January, 1989 at the United States Department of Education's Office of Educational Research and Improvement, where eighty participants from forty-seven states analyzed the national status of evaluation in service programs funded by LSCA Title I, it became clear that ascribing resistance to lack of skill alone is too simplistic. Even when librarians are knowledgeable, they may not evaluate. Some of the conferees' reasons for abstinence had a philosophical basis, such as, "What we do can't be reduced to numbers"; others had an operational basis, such as, "Costs are too high and evaluation consumes more time than we have to give it." With some probing, however, two prevalent underlying reasons were brought forth. First, librarians have little faith in the usefulness of evaluations. For all of the effort assessment requires, they believe no one pays attention to the results. Second, all too frequently, evaluation militates against demonstrating the worth of nontraditional services for nontraditional populations. Taken together, these reasons pointed up the perceived lack of utility of evaluation, and the misinterpretation of evaluation as synonymous with currently practiced output measurement.

**Expanded Options**

In the last decade a shift has taken place in evaluation, from the dominance of numbers in quantitative assessments toward the addition of narratives in qualitative approaches. That shift is only now beginning to have an effect on library programs. Until twenty years ago, minimum standards for public libraries and public library systems issued by the Public Library Association (PLA) concentrated on the resources supplied to provide service, such as income, number of staff, volumes owned, and volumes added (Public Library Association, 1966). The major problem uncovered with these assessments was that putting standard inputs into a library did not necessarily assure standard levels of
activities, such as circulation or the number of reference questions answered per questions asked, i.e., input and service did not necessarily go hand in hand (Chelton, 1987, pp. 463-84).

In the 1970s, with a grant from the U.S. Department of Education, Ernest DeProspo at Rutgers University began building the case for support from a more systematically developed and tested set of quantitative measures that emphasized outputs, i.e., measuring performance through services used, such as library visits, in-library materials circulation, and program attendance (DeProspo et al., 1973). By 1982, PLA had sponsored the publication of *Output Measures for Public Libraries* (Zweizig & Rodger, 1982), which was revised in 1987 (Van House et al., 1987).

As adoption grew, problems were uncovered. Today, although output measurement may be managerially necessary, stressing it without regard for its limitations has retarded the development of library program evaluation, especially with regard to demonstrating the worth of services for special populations. Studies over time have revealed that when measures of use are compared, the differences discovered may not be due so much to service performance as they are to the social and educational characteristics of the library's public (D'Elia, 1980, pp. 410-30; D'Elia & Walsh, 1983, pp. 109-33; D'Elia & Walsh, 1985, pp. 3-30; D'Elia & Rodger, 1987, pp. 5-20). Even in the face of evidence that applying output measures may make difficult the fair assessment of services to special populations, particularly those situated in economically disadvantaged communities, they are still the only approach widely recommended.

The use of input and output measurement has also been called into question because it does not reflect on the quality of service provided. It makes no distinction between technical quality—what is delivered—and functional quality—how it is delivered (Shaughnessy, 1987, pp. 5-10). While currently outside librarianship the emphasis is on outcome measurement, we have failed to move into that realm even where it is most appropriate. The focus of output measurement is the library, but the focus of outcome measurement is the library's users. The shift is to determining impacts, that is, what happens as a consequence of a program. This approach takes a marketing rather than an institutional stance by asking such questions as: How well did the service meet the magnitude of the need uncovered? Did it have the intended effects? Did it reach the target audience? What changes occurred in them? Were their skills enhanced? Were they able to reach a personal goal which improved the quality of their lives or the lives of their family members? What values did they derive from library use?
The answers to these questions give a better picture of the merit of services for special populations than traditional measures such as circulation per capita.

Common constraints put boundaries on the course undertaken in all evaluations. The aim is to conduct a credible assessment for affordable costs within the available time. Staff expertise also determines the design implemented; it cannot be more intricate or complex than staff can handle. When design demands a level of skill that is not available, options include hiring consultants, giving staff short, intensive training courses, or isolating complex or difficult portions of the design for performance under contract (United States General Accounting Office, 1984, pp. 12-13). The self-diagnostic approach to library evaluation currently in vogue has led to librarians assuming the role of evaluator in addition to other roles demanded of them. Indeed, that not only requires time unavailable, but it may not be worthwhile in the long run. A study of the U.S. Department of Education’s National Diffusion Network (NDN), established to recognize and disseminate information and training on exemplary programs of educational innovation, has shown that most of the programs deemed outstanding were assessed by expert outside evaluators (Lynch, 1987, pp. 20-24). Librarians can stop the self-flagellation because they are not authorities in the craft of evaluation and realize that there are some things experts should be hired to do.

Measurement and measures have held the spotlight. But the application of measures alone does not ensure the systematic process that is a hallmark of rigorous evaluation. The demand for evidence that something good is happening can exert pressure to decide program merit on the basis of what is readily measured. This rush to quantify can damage progress in developing sound library programs for special populations aimed at long-term outcomes (Schorr, 1988). Ultimately, the way in which the evaluation of a program is conducted must be determined, not by the application of a few measures, but by the questions it seeks to answer, the model that will best supply the answers, and the design that will uncover an accurate reflection of the program under scrutiny. In some cases, qualitative data is needed first to better understand and measure what will adequately assess impact, particularly where services to special populations are concerned. But qualitative evaluation is rarely discussed and even more rarely implemented.

Two Perspectives on Rigor

Qualitative strategies frequently supply the only means to fairly and accurately assess what is occurring in services aimed at special populations. Perhaps they have largely been ignored because they are
mistaken for a return to the conventional wisdom or because their rigor is questioned. But neither quantitative nor qualitative evaluations has a corner on rigor. They seek to answer different questions.

Qualitative strategies are directed toward descriptive questions. Quantitative strategies are directed toward normative and cause-and-effect questions (United States General Accounting Office, 1984, pp. 1-2). Descriptive questions provide data on the condition of program participants, why they need the program undertaken, how to reach them and provide them with service. For example, an English-as-a-second language program for older adults will have limited access to previously gathered systematic data to guide program implementation. The first evaluative step, then, is to collect information that will lead to an understanding of what is going on in the lives of the elders and how that will affect the way in which the service is designed and delivered.

Normative questions provide data that compare what is observed to what was expected, a standard of performance, or a performance objective. For example, the influence of a homework hotline for disadvantaged youths may have been discovered by comparing scores on high school assignments before and after program participation. As the number of scores mounts up over time, the program will develop a standard for improvement by which continued program success can be measured and by which the effectiveness of this program can be compared to other similar programs. Cause-and-effect questions collect data that reveal whether an observed result can be attributed to the program's operation, for example, determining what part of the change observed in the quality of research papers submitted by disadvantaged high school students is attributable to the effects of the public library user instruction program they attended. The proof may be determined by comparing a group who participated in the program with a group who did not.

That is not to say quantitative strategies should be cast aside. Michael Quinn Patton (1987) has created a series of questions to guide the determination of the appropriate approach. Quantitative strategies are preferred when:

1. Standards exist by which to judge the merit and worth of a program.
2. Program goals are specific and measurable.
3. Concentration is on comparing participants of the program on standardized, uniform measures.
4. Instruments are available to measure important program results.
5. Instruments can be developed that measure important results.
6. Emphasis is on aggregating information so that uniformities are highlighted.
7. Causes of change in the target audience are the focus of the evaluation.
8. It is necessary to apply statistical tests of significance to the data.
9. Information is needed on the generalizability of the program's results.

But qualitative strategies are preferred when:
1. The evaluation will assist in developing standards where none currently exist.
2. The evaluation is intended for a new, innovative, or demonstration program.
3. No valid, reliable, and believable instruments are available or readily capable of being developed.
4. The program is at the formative evaluation stage, where goals and program content are still being developed.
5. The goals of the program are vague, general, and nonspecific.
6. The focus is on diversity among program participants or events, and their uniqueness.
7. Detailed, in-depth information is needed about unusual failures or other critically important instances for financial or political reasons.
8. Information is sought about the details of program implementation, such as what participants in the program experience, what services are provided, how the program is organized, what staff do, what is going on in the program, and how it has developed.
9. Descriptive information is needed about the quality of program activities.
10. It is possible that the program is affecting participants in unanticipated ways (pp. 41-42).

Figure 1 compares the ingredients set forth by Yvonne S. Lincoln and Egon G. Guba for a rigorous evaluation under the two strategies (1985, pp. 294-301).

<table>
<thead>
<tr>
<th>Common Terms</th>
<th>Qualitative Evaluation</th>
<th>Quantitative Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truth Value</td>
<td>Credibility</td>
<td>Internal Validity</td>
</tr>
<tr>
<td>Applicability</td>
<td>Transferability</td>
<td>External Validity</td>
</tr>
<tr>
<td>Consistency</td>
<td>Dependability</td>
<td>Reliability</td>
</tr>
<tr>
<td>Neutrality</td>
<td>Confirmability</td>
<td>Objectivity</td>
</tr>
</tbody>
</table>

Figure 1. Components of a rigorous evaluation

**Truth Value**

Quantitative strategies approach truth by safeguarding internal validity. When an evaluation has internal validity, change caused by
the program can be distinguished from change resulting from other factors. Threats are avoided through controlling or randomizing sources of confusion.

Qualitative assessments approach truth through a determination of credibility, not a determination of causality. To establish credibility, the qualitative evaluator: (1) has extended contact with the program; (2) establishes review of the evaluation record as it is being created by a disinterested peer; (3) performs an active search for negative instances that may add insights to developing explanations; and (4) sets up checks during and at the close of the evaluation by a representative group of stakeholders to see if the reality which it presents is one that they agree represents the program.

Quantitative assessments approach applicability by safeguarding external validity. When an evaluation has external validity, the findings are generalizable, which is particularly important when results from current program participants will be used to make decisions affecting future participants, or when results are going to be applied elsewhere. Quantitative strategies ward off threats through random sampling which produces representative participants and allows precise statements about external validity. Within given confidence limits, the findings from the sample are considered to hold for the population represented. The results are said to extend to all environmental contexts within that population; they are generalizable.

Qualitative evaluators point out that the criteria of internal and external validity are in a trade-off situation by their definitions. If, for control, strenuous laboratory-like conditions are imposed on evaluations, then their results are not generalizable except in situations like the original laboratory. Threats to internal and external validity are a natural state of affairs for the qualitative evaluator, who must address them in making judgments of transferability. Here the evaluation sets out results with a description of the time and context in which they were found to hold. To be sure that the program and its success will transfer to other sites, it is not enough to know about the situation of the original program. Knowledge of the context to which it will be applied, and its similarity, is equally as important (Lincoln & Guba, 1985, p. 316).

Consistency

Quantitative assessments approach consistency by safeguarding reliability. The cornerstone on which reliability is built is replication. When an evaluation has consistency, two or more repetitions of essentially the same program under essentially similar conditions will yield similar findings. Qualitative assessments substitute proof of dependability for reliability. What happens in a program often varies
over time because of changes in the program, or because of changes in participants or changes in the emergent design of the evaluation as insights grow.

To demonstrate dependability, the qualitative evaluation relies on the external audit. Detailed records are kept during the evaluation of process, procedures, and evaluator insights, which establish an audit trail. Then review of the record is carried out by a competent external, disinterested auditor or second evaluator. If an evaluation is dependable, the auditor's findings will agree with the original evaluator's.

**Neutrality**

Quantitative assessments approach neutrality by safeguarding objectivity. They attend to the question of the degree to which findings of an evaluation are determined by the participants and the conditions of the evaluation and not by the biases, motivations, interests, or perspectives of the evaluator. To avoid this bias, the quantitative evaluation relies on detailed design before the evaluation begins. Insulation of the evaluator is equally important to objectivity, since it is easy to be influenced by what is learned, and that is considered damaging.

Qualitative assessments establish neutrality through confirmability. The control device is agreement by multiple peers on findings as expressed by the program evaluation. The qualitative evaluation proceeds from the assumption that the evaluator cannot maintain an objective distance from the program being studied; rather, the relationship is one of mutual and simultaneous influence. Far from being value-free, all evaluations are value-bound.

**Authenticity and Trustworthiness**

To summarize the differences between the two approaches to evaluation: The qualitative approach is built on flexibility in deciding what data to collect, from whom, and under what circumstances, and in organizing the evaluation according to the meaning of events to participants; whereas the quantitative approach requires having to decide beforehand on a set of data elements or on an essentially immutable plan of action. Qualitative assessments seek understanding of the local situation, while quantitative assessments seek to prove that a program successful in one library would benefit other locations.

In practice, the two approaches are frequently combined. Indeed, there is often a flow from one to the other. After the exploratory work of finding out what the important questions are, completed in qualitative phases, the evaluations of similar programs may switch to quantitative
testing aimed at confirming causality and then return to qualitative strategies to look for rival assumptions and unanticipated or unmeasured factors that may be influencing results.

Eight Models for Assessment

After the evaluation questions and strategy are decided, the model for assessment is selected. To date, the evaluation of programs for special populations has, in the main, relied on goals and objectives. But at least seven other approaches have been identified which can satisfy the underpinnings for rigorous evaluation and provide the trustworthy results that until now have eluded us in librarianship (House, 1978, pp. 4-12; & House, 1980, pp. 4-12, 21-43).

Quantitative strategies are represented in four models and qualitative strategies in an additional four. The Decision-making, Systems, Goal-Based and Goal-Free Models are all quantitative.

Decision-making Model

When utility is a hallmark of evaluation, program assessment is imbued with the Decision-making Model. The process is initiated by identifying stakeholders who have a share or an interest in the program under study from relevant constituencies and organizing them for input into the conduct of the evaluation. Three primary means that serve this purpose are stakeholder interviews, focus groups, and community forums (Rossi & Freeman, 1985, pp. 124-30). All provide an economical means of information gathering while developing support from community influencers.

For the evaluation of services for special populations, it is especially important to ensure that the stakeholders selected are: (1) knowledgeable about the community, its people, their needs, and the patterns of services already being delivered; (2) recognized leaders who are accessible; (3) representatives of the program’s target population; and (4) consumers of the program in addition to program designers and staff. Stakeholder check sessions are built in so that judgments of the overall credibility of the evaluation, statements of major concerns and issues, and statements about factual or interpretation errors can be identified.

How does this model apply to services for special populations? Decision-makers should be part of every library program evaluation. For example, at the close of federal funding for an information and referral service targeted to older adults, the board of trustees will decide whether or not to continue the service initiated by a grant under LSCA Title I. At the same time, the president of the board of trustees wants
a political career and one criterion affecting his or her decisions about library programs is whether or not they will increase visibility in a positive way among possible future constituents.

Interviews with board members and other stakeholders will form the basis for designing an evaluation that speaks to the information needed for decision-making. It is important to get below the surface and determine the real information sought. In evaluating the program serving older adults, information about the number of voters among elder participants, for example, would be as important as information about the number of elders who take part in the program.

**The Systems Model**

Typical questions addressed by the Systems Model include, “What impact did this program have? Can the results be produced more economically?” Library program evaluators who use this model collect data on a few well-defined outcome indicators deemed critical, for example, the per capita ratio of Information and Referral questions answered directly and by telephone to the total older adult target population. Variations in the measures are associated with differences in program outcomes, such as the improved ability of older adults to locate appropriate health caregivers. Generally, higher scores on measures are interpreted as meaning greater success. The relationship of outcome measures to program achievement is demonstrated via statistical techniques. The programs determined most effective have the highest possible activity measurement at the lowest possible cost. Many of these evaluations use test scores as the only measure of success. They are compared to normative data gathered on large numbers of similar cases over an extended period of time.

**Application to Services for Special Populations:** The Systems Model is appropriate for program evaluations that can compare participants’ pre-program and post-program scores to standardized scores, empirically demonstrating the extent of the program’s effects. One of the programs for special populations to which this model could be applied is literacy. Since there are numerous valid, reliable, standardized tests of reading achievement, *before* and *after* scores for literacy program participants provide strong evidence of program effectiveness. Unit cost measurement is added to demonstrate program efficiency.

For example, a library introducing two new methods of literacy tutoring might want to determine if one made more of a contribution than another to reading ability. Three groups of participants would be established and tested with standardized reading achievement tests before the new tutoring methods were begun. Then two of the three groups would be assigned to one of the two new methods; the third would continue with the earlier method. At the end of the program’s
funding, comparisons would be made among the achievement test scores of the three groups to determine whether there were significant differences in reading ability. The cost of the materials could be divided by the number of clients who used them, or the number of times each was used, for a unit cost figure.

**Goal-Based Model**

The most familiar approach and the most popular among evaluators, this model is also currently the most commonly advanced idea for evaluation. The primary question of the Goal-Based Model remains, "Is this program achieving what it intended?" Here, the identifying feature is the presence of goals and objectives. The object is to collect evidence to determine whether the program has achieved what it stated it would. The goals and objectives are the criteria by which the evaluator assesses what the program accomplished against what its developers started out to do. The discrepancy between the stated goals and the program's results is considered the measure of program success.

Proponents stress the accountability aspects of the model, since the program claims were the basis upon which the effort was mounted. Not unexpectedly, the Goal-Based Model has supplied most of the framework for the contemporary evaluation of public library performance. The extension course, "Are We There Yet?," developed by Jane Robbins and Douglas Zweizig, provides a step-by-step approach to the implementation of this model (1985, pp. 624-27).

**Application to Services for Special Populations:** The Goal-Based Model is a natural candidate for the evaluation of services for special populations. For example, a program might have as its goal improving services to the physically handicapped. An objective might be to locate and survey the needs of 10 percent of the physically handicapped population in the library's service areas in the first six months of operating a new Media Home Delivery Service. As one measure of success, the evaluation might compare the percentage located and the percentage surveyed against the target.

**Goal-Free Model**

Created in direct reaction to the ubiquity of the goal-determined evaluation, the Goal-Free Model was developed to reduce bias. It requires an outside expert or an internal evaluator unconnected to the program under review to carry it out. The major question it addresses is, "What are the intended and unintended effects of this program on its participants?" The evaluation is not based on program goals. In fact, the evaluator remains uninformed about them and searches for all
program outcomes, many of which are side-effects or unintended results, both positive and negative. In this case, it is not intention that is sought, it is achievement.

Among the models presented to this point, the traditional notion of objectivity has been built on quantitative assessment alone, but the goal-free notion of objectivity developed first in the qualitative realm. It can combine both strategies. Consumers Union uses this model in focusing on product criteria that it thinks will benefit consumers.

Application to Services for Special Populations: The Goal-Free Model would be applicable to many types of programs for special audiences. For example, a program funded to provide materials to support after-school reading is meant to increase skills in the reading disabled by exposure to a wide range of high interest, low reading ability materials. A number of qualitative and quantitative indicators might point to the success of the program. Examining the pre-program and post-program test scores of the students, visiting the scheduled tutoring sessions, interviewing tutors and students, reading expert reviews, and examining the materials themselves would provide abundant data that could substantiate success or failure.

Qualitative strategies are represented in four models for assessment. They include the Art Criticism, Professional Review, Judicial, and Case Study Models.

Art Criticism Model

This approach relies on critical review, the major assessment tool of the arts. Evaluators draw on their own experiences and intuitive reasoning to judge what is happening in a program and to express their judgments in a way that nonexperts can understand. Some questions that the Art Criticism Model seeks to answer include: "Would an expert approve this program? Are the people for whom the program was designed being helped? Are they acquiring habits conducive to their further development?"

Like an art critic, the evaluator, who is an expert in the program's speciality, uses the critical review to render the essential qualities of the program and make judgements based on her or his own standards of excellence. The critic-evaluator presents feelings as well as facts about the program. Proper training and experience are necessary to make evaluative discriminations; the evaluator must have both in sufficient measure to be able to distinguish what is significant. The evaluative report will heighten the awareness of its readers as to what constitutes a good program and so improve future program standards.

Critical review is accomplished in a couple of fairly standard ways. Immersion in the program is vital. Notes, video tapes, and similar recording devices are used to retain observations and the qualitative
procedure called Referential Adequacy is invoked. A portion of the data collected is archived and not included in the initial analysis. Later, it serves as a benchmark for comparison against a follow-up data analysis and interpretation to determine if features to which the critic pointed can be found in the archived data. The second data review also demonstrates whether different analyses reach similar conclusions.

Application to Services for Special Populations: This model would provide a good option for application to the evaluation of library programs for latchkey children. An evaluator using it would have been immersed in problems in the lives of latchkey children as well as in services that respond to those problems. She or he would be familiar with library programs considered exemplary across the country and the elements that led to success. The review of the specific program and the judgments expressed in the evaluative report would inform and educate those evaluated and/or less knowledgeable. The critical review would be based on extended observation, continuing over a period of at least a month. The narrative would establish the strengths and weaknesses of the program, offer comparisons to exemplary programs that might exist elsewhere, and make recommendations for improvement.

Professional Review Model

Conducted by a team of peers who have the qualifications to judge the merit of a program, this model culminates in a holistic assessment by other professionals (Dressel, 1971, pp. 277-87).

Before evaluators visit the site, the staff engages in self-evaluation. They are appointed to committees that review each of the program's functions and prepare a program profile. When turned over to the peer reviewing team, the self-study includes: definition and clarification of program purposes and goals; examination of the adequacy of resources; an appraisal of the quality and morale of the program staff; a review of the strengths and weaknesses of the current organization and delivery methods; consideration of the overall program climate and environment, including the role of clients and their satisfactions and dissatisfactions with the program and its services; and finally, a collection of evidence on the effectiveness of the program and the process of client development. Before they leave, in their evaluation members of the peer review panel indicate their differences from the staff review, give a brief oral report pointing out the strengths and weaknesses of the program, and make recommendations for change. After the visit, the program is expected to correct perceived weaknesses.

Application to Services for Special Populations: Using the Professional Review Model to evaluate an adult basic education program, one of the criteria established to determine excellence might be that
“attention is given to improving study skills.” The review panel, using a checklist, might find that item and mark the quality they believe existed on a five-point continuum from missing to excellent. Each of the major program functions would have similar checklists where criteria would be evaluated. The checklists would be totaled for a holistic appraisal of the program.

Judicial Model

Blue Ribbon Panels, like the Kerner Commission or the Warren Commission, fall within this approach. Presidentially appointed, members of these Panels heard evidence from witnesses, conducted their own investigations, and came to conclusions about probable occurrences in two momentous events in history.

The Judicial Model is based on the supposition that the facts in a case are uncovered best if each side strives as hard as it can, in partisan fashion, to bring the most favorable evidence for its view to the attention of the panel. The aim is to resolve the issue of how a program should develop in the future. Evidence is presented to demonstrate the program’s strengths and weaknesses. The approach is patterned after the courtroom. Rules are formulated about who may testify and the conditions for testimony. Evidence includes not only facts, but also feelings, perceptions, opinions, biases, and speculations. The Judicial Model has four stages: issue generation, where sometimes as many as thirty or more interviews are conducted; issue selection, where surveys are undertaken to hone in on what is crucial; argument preparation; and a hearing. The major advantage of this model is that pressing issues can be addressed quickly by the panel who bring about an immediate resolution to future directions.

Application to Services for Special Populations: Clearly, the approach has promise for programs which may need revamping in midstream. For example, in a decision about whether or not to continue to fund the public programming elements of a library-based career center in a community where unemployment is high, members of the Blue Ribbon Panel, appointed perhaps by the State Library, would interview key members of the staff to ferret out the issues. To gather opinions, they would develop a questionnaire and send it to a broad number of stakeholders including administrators, persons served, and government officials, in addition to staff members. Arguments would be prepared for and against the continuation based on that data and the opinions of partisans. A hearing would take place before the panel and a decision would be made by the members following the hearing’s conclusion.
Case Study Model

The final qualitative model provides a way of judging programs within the context of their environment. Rather than pushing for quantification, this model pushes for understanding. Its strength lies in its ability to assist us in determining how to create programs that are responsive to nontraditional audiences. Here stakeholders observe the program and assist in its evaluation.

Evaluators report on the perceptions of others as well as their own in giving their judgment of a program. Since this model attempts to improve the understanding of the audience, the program staff, and sponsoring agencies about the program is and what is going on in it, the aim is to collect data to demonstrate how the program is perceived by others, particularly by the audience it was intended to serve.

The case study is usually reported as a narrative with a great many quotes directly from the participants' own words. Actual instances are cited and observation is the primary data collection technique; it substitutes more objective experiences for anecdotes of unknown credibility (United States General Accounting Office, 1987, p. 59). This model concentrates on the description of program processes as well as outcomes. Program observers prepare and submit narratives, portrayals, and graphics to stakeholders for feedback. Evaluators find out what is of value to program audiences and gather expressions of worth from various individuals whose viewpoints differ. They check the quality of the records, get program personnel to react to the accuracy of their portrayals, and get stakeholders to react to the relevance of the findings.

Application to Services for Special Populations: There is no approach that gives better results for the evaluation of new or innovative programs than the Case Study. For example, an application might be to a program for high school dropouts that intends to provide nontraditional means to earn a high school diploma. Since the library has had little systematically evaluated experience in this area, the Case Study could bring a better understanding of what is needed to make such programs successful and to provide for their transportability to other library locations. In addition to gathering perceptions of program strengths and weaknesses, the study would provide extensive description of the context in which the program was conducted and how that context affected daily operations.

Although the models are separated into quantitative and qualitative strategies here, their actual differences are often not so cut-and-dry. Combinations frequently provide the basis for the best-case scenario to prove library programs for special populations work. They can and should be mixed and matched to meet the needs of the evaluation. Numbers can add authority to the Case Study; narratives create the
context that adds authenticity to numbers. The combinations do not dilute the validity of the process as long as systematic procedures are followed in creating and implementing the evaluation design.

Rigorous evaluation shows a link among the major components of the evaluative process—questions, strategies, models, and measurement. The determination of the measures on which to collect data does not precede the process; it is a result of it.

Measuring Results

Input, output, impact, and cost measures are all useful when evaluating the worth and merit of a program of service to special populations. Figure 2 compares the definition, purpose, and elements on which program-related data are gathered for each of these measurements. Output Measures for Public Libraries, second edition (Van House et al., 1987) and Cost Finding for Public Libraries: A Manager’s Handbook (Rosenberg, 1985) supply data collection techniques for output, and costs that can be adapted to evaluation. Evaluation of Adult Literacy Programs (Zweizig et al., 1990, pp. 39, 42) provides a few measures of impact which are amplified here. Once again, a most persuasive case can be made by creating combinations, this time of measures.

For example, in a community where no high school diploma is granted to students who read below the eighth grade level, the library set up a “ Teens Top the Mark” program in cooperation with the local school system. In the application for LSCA funding, the problem statement clearly denoted the target population. Out of an annual graduating class of 400, about 10 percent failed to receive diplomas based on their inability to read at the appropriate level; that number had increased in each of the last five years. In the past, these students had not experienced success in traditional remedial reading classes established to help them improve their skills and graduate.

The “ Teens Top the Mark” program was introduced by converting a little-used branch into a tutoring and homework facility staffed by teacher-librarians and stocked with young adult materials. The library’s program incorporated a new approach modeled after adult literacy programs with confidential one-to-one tutoring. The tutors were volunteers who themselves learned to read proficiently as adults. All students who, at the beginning of their junior year, are in danger of not graduating because of lack of reading skills were recommended to the program.

The evaluation employed an interrupted time series design. Measurements were taken before and after participation in the program. Scores were recorded on a standardized reading test to show the impact
on skills. A questionnaire captured data demonstrating the impact of the program on attitude and behavior related to reading, library use, and the program participants' views of themselves as self-learners. The questionnaire also measured participants' satisfaction with the quality of the program and facilities. Records of library use were kept for each student. Input and output data were gathered on the resources allocated to young adult services and on overall library use.

At the end of the year-long program, students had achieved an additional three years as determined by scores on a standardized reading achievement test taken before and after they participated in the program. They were no longer held back from reaching their personal goal of obtaining a high school diploma. The intent of the program was also met because the high school accepted the tutoring program as a valid means of gaining the level of competence needed, even though it did not contain all the elements prescribed by the high school's own remedial reading program. Attitudes on library use and reading showed significant improvement. Of the target population's forty students, thirty-five were eligible to graduate, 50 percent more than in previous years under other programs of remediation. The federally funded program had attained its intended impact.

Figures on output measures gathered one year after the program's initiation also showed that, for the target population, library visits quadrupled, the number of library cards issued had doubled, and circulation was three times larger. Input data documented that the library's expenditures for young adult programs from its locally supported budget had also doubled. When the per capita costs of running the seldom-used branch were compared to the per capita costs of running the branch once the program was up and running, a 25 percent decrease was calculated. At the time of graduation, nine months after the program's conclusion, there was no deterioration in reading skills. The proof of worth and merit was made.

CONCLUSION

The fact that evaluation results have led to so few action agendas is virtually a national scandal. A posture that includes stakeholders and empowers them to change the decision-making process holds promise for eliminating that lack of attention.

Diversity in design is incorporated into the models of evaluation recommended. While the Goal-Based Model currently embraced is worthy of consideration, it is not the only approach for evaluation to take. We have swung from assessment based on the conventional, collective wisdom to quantitative measurement without recognizing the
Population Measures
Definition: Potential and actual number of program participants
Purpose: Demonstrate the program reached its intended audience
Gather Program Related Data On:
- Total Population in Service Area
- Number of Potential Program Participants
- Ratio of Potential Participants to Total Population
- Number of Actual Program Participants
- Ratio of Actual to Potential Participants
- Number of Program Participants Reaching Program Standard for Success

Input Measures
Definition: Resources allocated to support a program
Purpose: Demonstrate improve institutional practice
Gather Program Related Data On:
- Income
  - Local Taxes
  - Capital Income
  - Federal Funds
  - State Funds
  - Endowments
  - Foundations
- Expenditures
  - Salaries and Wages
  - Materials
  - Per Capita Expenditures
  - Capital Expenditures
- Staff
  - Librarians
  - Volunteers
  - Others
  - Full Time Equivalents (FTE)
- Materials
  - Owned
  - Purchased During Program
- Facilities
  - Square Feet of Building Space
  - Number of Buildings or Sites

Output Measures
Definition: Performance on services emanating from a program
Purpose: Demonstrate improved institutional support
Gather Program Related Data On:
- Circulation
- Turnover Rate
- In-library Use of Materials
- Library Visits
- Number of Library Cards Added
- Reference Transactions
- Attendance at Programming

Figure 2. Selected population, input, output, impact, and cost measures (cont.)
Impact Measures
Definition: Outcome or Consequences of a Program
Purpose: Demonstrate Enhanced Skills and Changes in Attitude and/or Behavior
Gather Program Related Data On:
Enhanced Skills
Behavior
Time Spent Reading
Comfortable Use of Other Libraries
Increased Visits to the Library
Borrowing More Materials from the Library
Attitude
Desire to Read
Improved View of Self as Learner
Attitude Toward Reading Improved
Satisfaction with Program
Satisfaction with Program Facilities
Perceived Match Between Program Expectations and Experience
Achievement of Personal Goals

Cost Measures
Definition: Funding Required to Finance a Program or its Components
Purpose: Demonstrate Improved Institutional Practice
Gather Program Related Data On:
Unit Cost, the Cost of Supplying One Unit of Service
Cost Per Capita, the Cost of Supplying One Unit of Service to One Program Participant

Figure 2 (cont.). Selected population, input, output, impact, and cost measures

many approaches available. The model pursued should fit the environment in which the evaluation is being conducted, mesh with the purpose and situation under assessment, and retain the rigor necessary for it to command the respect of evaluation experts. Given the constraints under which library programs operate and the little systematic evaluation undertaken, multiple models must be introduced and encouraged.

Since bad evaluations can irreparably damage programs and injure the constituents for whom they are intended, they must take into account more than measurement and measures. While in the past the emphasis in public librarianship has been on the performance of the library, it is time to focus on the special populations for whom the programs of service were developed. In such a shift, the institution recedes into the background and the library user becomes the focus of attention. Without that reversal in perspective, evaluations cannot measure impact and programs cannot fulfill their public service missions.
REFERENCES