Moving from Counting to Programmatic Evaluation:
A Look at Best Practices

Claudia J. Dold and Ardis Hanson*

How do we teach librarians to assess and to evaluate effectively? Assessment and evaluation to most librarians means the listing and counting of things, events, and patron attitudes. The two processes—assessment and evaluation—are not equivalent. Assessment is the process of gathering and analyzing data. Evaluation is the process by which we determine the overall value of an outcome based on the assessment data.

Librarians are familiar with program assessment and evaluation. An accreditation committee reviews an academic department, in part, on its assessment of data provided by the department itself. However, the committee also bases its evaluation on other information gathered in site visits, impromptu interviews, measures against rubrics, etc. LIBQUAL is another programmatic assessment tool, which the University of South Florida has used in our libraries for several years in order to ascertain facets of library effectiveness (Cook and Maciel 2010).

The focus on this paper is on programmatic evaluation. The term “program” will be defined as “any set of organized activities supported by a set of resources to achieve a specific and intended result” (Centers for Disease Control & Prevention 2012). Programmatic evaluation essentially wants to know if the program activities were implemented as intended, did it achieve the objectives it was intended to accomplish, are the resources needed to conduct this program used efficiently, is the value of this program worth its cost, and is there a relationship between this program and its stated goals and objectives?

Since evaluation depends on assessment, to evaluate a program, one must decide what will be assessed at the program level. Classroom assessments, for example, are part of evaluation; however, evaluation is more than the sum of satisfaction surveys. The Kirkpatrick model, a worldwide standard for evaluating the effectiveness of training, envisions four levels to program evaluation (Lee and Pershing 2000). The lowest level is the reaction level: Did the participant like or dislike the program? The second level is the learning level: Did the participant acquire the knowledge that we taught? The next level is the behavior level: Did the participants change their behavior or thought process, because of the program? In an academic setting, the capstone course and paper is a measure of this transformation. The fourth level is the results level: Did the program do better than it did the last time it was offered? Each of these levels can be tracked by metrics that address program goals. The focus of this paper addresses the third and fourth levels of the Kirkpatrick model: Did the program make a difference in the way people see the world and interact with it? Did the program meet or exceed the institutional goals set by the library administration or the university?

Decisions made regarding the choice of assessment vehicles affect staff, services, resource allocation, and library patrons (Tatarka et al. 2010). Therefore, librarians need to be clear about their intended target of evaluation.

* Claudia J. Dold, MLS is Assistant Librarian at The Louis de la Parte Florida Mental Health Institute Library, University of South Florida, cdold@usf.edu; Ardis Hanson, PhD, MLS is Assistant Director for Research and Education at Shimberg Health Sciences Library, University of South Florida, hanson@health.usf.edu.
tion when they examine services, programs, or trainings: What change do they want to measure? They also need to determine accurately the organizational level of the assessment they intend to measure: Is it the individual, unit, or institutional level? Those choices will determine the input sources they will use in their evaluation. Assessment and evaluation become murky when these two major questions are confused. For example, when libraries assess information literacy outcomes, assessments often focus on student learning, not on the effectiveness of the program (Ayre et al. 2015, Bailey and Paul 2012, Bowles-Terry 2012, Murray 2011). As Pogo said, we have met the enemy and he is us (Kelly 1970).

Given the current emphasis on evidence-based practices in librarianship, we librarians also need to focus on evidence-based evaluation. That requires a significant commitment to the **intent** of evaluation as well as the **process** of evaluation. Unfortunately, many library evaluations fail the evidence test. We look for best practices but there is seldom a guide to tell us how a specific program will be designed, implemented, and evaluated. Because we have few or no practice standards, our programs lack fidelity to implementation. Because we had no clearly stated research questions about our program when we started it, assessments tend to be self-reports and satisfaction measures. Our reported outcomes may be difficult to verify or even to correlate with our programs. While affective measures, such as user satisfaction, can affect behaviors and skills development, they cannot replace more reliable measures of skills and knowledge (Schilling and Applegate 2012). Hence, a major challenge for librarians is to learn how to balance non-measures (observations), indirect measures (self-reports), and direct measures (tests). Further, quantitative analyses can elicit and illustrate unique relationships among variables for assessing learning, such as efficacy of tests, attitudes, and reported behavior changes that are involved in assessing and improving programmatic and student outcomes. With the increased emphasis on assessment and data literacy, librarians should know the different kinds of data, which data to use for evaluation, as well as to generate, interpret and apply assessment data. This allows librarians to use evidence (data) to inform practice.

**Rethinking the Design of a Program Evaluation**

Program evaluation is a series of activities that contribute to program or project management. Effective program evaluation has five components. It addresses needs assessments, monitors program activities, measures program outcomes, informs decisions about program improvement or continuation, and helps plan future projects. In short, the evaluation should be designed to determine whether the program works and produce evidence about how one knows it works.

| TABLE 1 |
| Evidence-based Evaluation |
|---|---|---|---|---|
| Identify the Program Design | Ensure Effective Full Implementation | Assess Program Outcomes | Obtain Evidence of Positive Program Outcomes | Attain Strong Evidence of Positive Program Outcomes |
| Develop logic model | Fidelity | Process/Outcome evaluation | Multiple pre- and post- evaluations | Multiple evaluations (quasi-experimental design) |
| Pilot implementation | Process evaluation | Outcome evaluation | Impact evaluation |
| Evidence-informed | Evidence-based |

Two types of evaluation, formative and summative evaluations, are commonly used in program evaluation; however, process, outcome, and impact evaluation also play a role in determining the larger scope of program evaluation. **Formative evaluation** ensures that a program or program activity is feasible, appropriate, and ac-
ceptable before a program is implemented fully. It is usually conducted when a new program or activity is being developed or when an existing one is being adapted or modified. Summative evaluation occurs at the end of a designated time-period and is effective for review of long-term progress on major program goals and objectives.

Process evaluation determines whether program activities have been implemented as intended. Process evaluation helps to inform changes or improvements in the operation of a program. It documents what the program is doing, to what extent it is doing these activities and how consistently the program has been doing these activities since it was implemented. Process evaluation includes both qualitative and quantitative information and does not require a comparison group. Outcome/effectiveness evaluation measures program effects in the target population by assessing the progress in the outcomes or outcome objectives that the program is to achieve. Impact evaluation assesses program effectiveness in achieving its ultimate goals. Impact evaluation requires a comparison group and is generally quantitative. These last three evaluation facets sit at the fourth level of Kirkpatrick’s model.

Another important concept in evaluation is fidelity, which can be defined as “the degree to which … programs are implemented … as intended by the program developers” (Dusenbury et al. 2003). In short, did we do what we said we were going to do? An alternative definition that is applicable for library programs is ‘organizational members’ level of commitment to using the distinct components that make up an intervention as they were designed’. Both definitions reflect individual and organizational commitments to standards of practice that are measurable, observable, and able to be evaluated. Given an understanding of the measures of program evaluation, the next sections address the importance of developing a viable scope of work, clarifying research questions, developing a logic model, and fidelity of implementation.

Developing a Scope of Work
The first step in program evaluation is to determine the scope of work that the evaluation will comprise. Essentially, program evaluation planning begins as the project is designed. Project milestones, reports, deliverables, and a projected timeline for all work phases are contained in a scope of work. If these terms do not seem applicable to library work, then consider an approach that integrates service management (responsiveness to clientele) and operations management (improvement in efficiency), which examines various components of program evidence, efficient operation, client satisfaction, and positive outcomes. As F. W. Lancaster (1988) pointed out almost 30 years ago, program evaluations are an essential component of management and administration. From a business perspective, program evaluation 1) establishes a benchmark to measure the effects of a program’s change on its performance, 2) compares service performances across libraries, 3) justifies the existence of a program through a service benefits analysis or a cost-benefit analysis, or 4) identifies potential or actual shortcomings, inefficiencies, or need for change. To conduct a program evaluation requires an understanding of the scope of the program and, as mentioned before, its intent. Both of these are contained in a scope of work.

A good scope of work will include a number of tasks. These include 1) identifying the activity or strategy to be evaluated; 2) stating the purpose, the audience and use of the evaluation; 3) providing a brief background of the need for the program at the time it was planned; 4) identifying existing performance information sources; 5) clarifying the evaluation questions; and 6) identifying the evaluation method to answer those questions. Once this has been done, there are three additional steps that need to addressed: consider the composition of the evaluation team and identify the program stakeholders; review the planned and implemented procedures, such as schedule, logistics, and budget; and finally, clarify the requirements for reporting and dissemination. In essence, one creates and follows a comprehensive template.

While these lists may appear extensive in order to evaluate or assess an activity, it is essential for the scope of work to provide the same foundational knowledge to all members of an evaluation team and to subsequent
teams that may want historical information on the why, what, and wherefores of a program. It also means everyone who will be involved with the project, whether an active team member or an administrative reviewer, shares an understanding of the project.

We advocate for a discrete project management approach to programmatic evaluation. Auditing follows a similar process. The auditor decides what area of a business (s)he wants to examine, how to will examine it, what tests to perform, and what measure of results will be sufficient to make a conclusion. The auditor always reviews the previous audit work papers and its final report. All the work is plotted as to sequence and time constraints, the tests are specified, the sampling technique is identified, and the shareholders and potentially interested parties are known. Frequently, there are contingencies. Suppose the audit reveals a problem with the cash account. Then other tests will be performed to identify the source of the leakage, and from practice guidelines and experience, those procedures are clearly defined and understood. Each audit is individualized to the task-at-hand, but the process is well understood. By starting with a focus on the information that the audit wants to identify, the remaining pieces fall in place, molded to the particular situation while the program architecture remains in place. Flowcharting of the work process is a useful tool to identify all the elements that interact with a process. Likewise, the Gantt chart is useful to track the progress of a project, and the Critical Path Method or Program Evaluation Review Technique (PERT) diagram marks critical deadlines and decision points. We believe that librarians can and should borrow tools that are used successfully in other disciplines to guide their evaluation work. The approach can be just as direct as a surgical intervention or a military mission. The target is known, the steps to get in and get out are clearly identified, and the most efficient resources and tools are brought to bear on the situation. One such tool is the ILIAC system for program evaluation in libraries (Oakleaf 2009).

By clearly identifying what is to be evaluated, one may avoid ‘mission creep’. We want to measure A; but the next thing we know, we are also measuring B, C, D, and something else, too. A good program evaluation avoids mission creep. We will know why we are evaluating this project and who will be the audience (administration, patrons, accrediting institutions, etc.). Historical knowledge is very important. The documentation explains why this project was started, when it was started, how it was started, and who started it. This knowledge allows the team to understand the context of the project at its origination point and the measures used to evaluate it. Too often, teams jump into evaluating a project without a full understanding and consideration of a program or project's historical significance or reasons for its implementation. Many times, the only available information on a previous project or program is a final report. Other documentation is missing or never written. Taking the time to create a thorough scope of work allows those individuals who follow up months or years later to conduct a full review of the intent of the project, identify the reasons for its evaluation for the current administration, and design future program/organizational reviews. Review teams will know which library staff comprised the original team and their respective roles and responsibilities. In addition, the scope of the work identifies external stakeholders, such as teaching/research faculty or library patrons, who will be participants in the evaluation process. Reports of the project/program should clearly identify how it was evaluated, indicate whom the audience was, and provide relative measures as well as suggested improvements or specific achievements that may still be valuable.

A prospective scope of work also addresses the evaluation schedule, logistical issues (such as setting up surveys, data gathering, and reminders to participate), setting milestones for specific activities to occur that may need to be sequenced, and time for analysis and synthesis. A scope of work helps determine what research questions will be asked and why these questions. The questions will guide the choice of evaluation (formative, process, outcomes, or impact). Finally, a scope of work clearly defines deliverables and who receives them: a report, a formal paper, presentation of the findings, a news release, or a memorandum to a vice-president.
The scope of work should also address the budget involved in conducting the evaluation. Ultimately, evaluation is costly, whether measured in time expended or real costs incurred. The work process needs to keep track of the number of people and the amount of time each one spent on the different components of the evaluation, in order to generate a cost-benefit report. At our university, for example, we often have to buy out time for certain staff, such as statisticians, for different projects, or we may have to buy a software license to use a statistical package.

Resources like these improve the evaluation, whether one is working with quantitative or qualitative data. Quantitative outcomes may include the increased number of students passing a qualifying exam with essay questions as a function of a library program that addressed writing an analysis and synthesis of the literature. Qualitative outcomes may include the determination of improvement in morale or a higher quality program as a function of an embedded librarian program in salient courses.

**Clarifying Research Questions**

Why is clarifying the research questions so important? First, the questions themselves are the foundation of a successful evaluation. They define the topics we investigate, in the same way that a hypothesis is the source of search terms. Ideally, the research questions precede the program implementation because they guide the evaluation planning process and provide structure to the evaluation activities. Given the research questions that define the evaluation’s purpose and scope, we can determine the type of evaluation design, e.g., process or outcome (Kirkpatrick’s levels three or four).

Once we know the topic of our research question, we may determine the information that the stakeholders need or hope to gain from the evaluation. For example, let us suppose that the evaluation is going to address institutional expectations, such as changes in mission, reaccreditation requirements, or supporting a long-term research agenda for your colleagues. We can now determine which components of the program are the strongest candidates for evaluation. There is no need to evaluate the entire program at once. By conducting formative and summative assessments, the librarians may choose to look at one program over five years, and evaluate different components of that program every year. At the end of five years, they will have an evaluation that may address process, outcomes, and impact, which ties back to the Scope of Work, which documents the reasons the librarians conducted the evaluation in the first place. Inevitably, an evaluation’s scope should align with resources. If there is only paraprofessional to assist with the evaluation in the library, the program evaluation plan needs to be designed to align realistically with staff, time, and money constraints.
Developing a Logic Model

To develop the research questions, we suggest creating a logic model that looks at process and outcomes. A logic model uses a theory-based approach to change and accountability. In addition, it provides a snapshot of how a program or project works and shows the relationships between different components of the program. A logic model also helps link evaluation instruments more closely to specific program objectives. Using a theory of change approach helps develop action-oriented strategies to achieve a program's intended goals, and creates explicit relationships among the populations the program will serve, the program's intended goals, and the strategies to achieve those goals (Hernandez and Hodges 2005). Since theories of change address the conceptualization of a project, how a project is operationalized, and how the project is actually implemented, the scope of work should contain the theoretical underpinnings of a project or program. As one may see in figure 2, inputs, activities, and outputs are processes, and changes are outcomes. Inputs are the tangible and easily quantifiable resources that are needed to run a program. Examples of inputs are numbers of staff, size of potential community to be served, and amount of expenditures. Since processes define how a program is run, activities are those actions, workflows, or decision paths that need to be conducted in order to convert inputs to outputs. One can also think of activities as comprising process improvement techniques. Outputs are the quantifiable units of service the program provides. However, quantified outputs do not measure the value of a program; outcomes are the impact that a program has on its users. In many cases, outcomes address change that occurs in library users (behavioral, knowledge, skills, etc.).

While logic models are common in social sciences and health systems research, service design is more common in businesses and libraries (Polaine, Løvlie, and Reason 2013). Each is designed to address ineffective assessment instruments, encourage iterative processes rather than linear processes, and avoid process silos, which result in a lack of coherence and cohesion in program development, assessment, and sustainability. Both also work at the strategic level, connecting institutions to the populations they serve. The benefit of incorporating service design into a logic model is that ‘blueprints’ of services show both frontstage and backstage elements of each process involved in an activity, which more fully informs inputs, activities, and outputs. These blueprints map user actions, physical elements (classroom, syllabus, library, staff, etc.), lines of interactions (each step involved in the activity), and internal interactions that occur backstage (behind the scenes).

Fidelity of Implementation

How individuals and organizations acquire “new” concepts, knowhow, or tacit mastery, across individual, social, systemic, and structural processes is a critical component of change. Generally, adoption and successful implementation of new programs are built upon acquisition of new knowledge processes and changes in user behaviors. Therefore, if a library decides to implement a new information literacy program (ILP) with six librarians teaching the program, determining the success of the program would examine how well the ILP was implemented with fidelity by all six librarians in all the classes they taught.

Part of Rogers' diffusion of innovation theory (Rogers 2003), fidelity of implementation is used to understand how new ideas are put into practice. Since organizational culture, structure, climate, and work attitudes
may affect fidelity of implementation, procedural development of a standard curriculum is generally one of establishing a framework for accepted execution of tasks, a formal identification of task process, or an outline for operating requirements, policy or modus operandi (Rogers, 2003).

For example, if each of the six librarians are given the new ACRL information literacy guidelines as the starting point, there may be six radically different classes, assignments, outcomes, measures, etc. developed. In this case, it would be impossible to determine if the six programs were effective in information literacy as a programmatic initiative. However, if a standard curriculum and assignments were created, with allowances to tailor it to specific disciplines and databases, then one could measure how well the common elements of the curriculum and assignments met the identified KSA outcomes for the students and incorporate specific outcomes unique to a specific discipline. Hence, we are looking at both the context and the content of the ILP and the teaching skills/methods of the librarians. While qualitative methodology and survey/questionnaire methodology are the predominant measures used to examine diffusion, adoption, and implementation, formative, summative, and comparative evaluations can elicit organizational, program, and user-oriented design and adoption issues.

**Designing Evaluation Research Questions**

Just like a well-designed research study, each evaluation should have well-defined research questions that play on the strengths of the type of evaluation being done. For process, we look at who, what, when, where, why, and how for each of these factors: inputs, program activities, outputs, and stakeholder views. For outcome, we look

<table>
<thead>
<tr>
<th><strong>TABLE 2</strong></th>
<th>Sample Logic Model for an Information Literacy Program</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Process-Based Evaluation</strong></td>
<td><strong>Outcome-Based Evaluation</strong></td>
</tr>
<tr>
<td><strong>Inputs</strong></td>
<td><strong>Activities</strong></td>
</tr>
<tr>
<td><strong>Funding</strong></td>
<td></td>
</tr>
<tr>
<td>Library staff</td>
<td>Develop scope of work;</td>
</tr>
<tr>
<td>Library faculty</td>
<td>Develop curriculum;</td>
</tr>
<tr>
<td>Teaching Faculty</td>
<td>Select assessment tools;</td>
</tr>
<tr>
<td></td>
<td>Faculty training (library and teaching faculty); Develop interactive tutorials; Develop handouts; Develop webcasts; Reserve space for classes; Workshops; Publicity;</td>
</tr>
<tr>
<td></td>
<td>6 workshops/yr; 5 webcasts; 5 interactive tutorials; Trained 200 students</td>
</tr>
<tr>
<td></td>
<td>Trained 5 library faculty; Trained 10 teaching faculty</td>
</tr>
</tbody>
</table>
at changes, effects, and impacts across short, mid-, and long-term time periods, specifically in knowledge, skills, attitudes (KSAs), opinions, behaviors, actions, conditions, and status. For impact, we look at the changes on specific outcome factors for a specific group compared to another specific group, such as English class A versus Engineering class B. Consider the three questions listed below [and the terms in brackets] as a simple way to remember the differences between process, outcome, and impact evaluations.

- **Process**: Who, what, where, when, why, how is the program, model, component for evaluation purpose?
- **Outcome**: Did model, program, program component have a change, effect on outcome(s) for individuals, groups, or organizations?
- **Impact**: Did model, program, program component have an impact on outcome(s) for individuals, groups, or organizations relative to a comparison group?

All research evaluation questions must be clear, specific, and well-defined. Generally, comparing apples with apples and oranges with oranges is easier than comparing rocks with starfish. If you want to compare rocks with starfish, clarity is essential to conduct a valid evaluation. Keep a clear focus on what you want to evaluate. The logic model, which delineates inputs, activities, and outputs will help identify those elements you really need to examine.

Make sure your questions are designed to fit the evaluation. When the question does not correspond to the evaluation or data is “massaged” to answer an “unasked” question, the trustworthiness of the evaluation is diminished. For example, when one wants to evaluate the transfer of knowledge within a program, a question about how well the participants liked the program may not be effective. Finally, make sure your questions align with your logic model. Again, if the question is not documented as part of the process (input, activity, or output) or outcome, then what are you really trying to evaluate?

**Concluding Thoughts**

As Wellman (2000, September 22) reminds us, “assessment of learning is an imperfect science, one that has not yet evolved into measures that are commonly understood and easily transferable to different types of institutions” (para. 10). The assessment and evaluation of learning uses social sciences tools that allow librarians to interpret data, determine relationships, and make causal and/or correlational inferences for what is deemed ‘successful’ or ‘unsuccessful’ in our attempts to define best practices. However, programmatic evaluations take time and resources to determine those educational outcomes that adequately assess behavioral changes and knowledge acquisition of library users. It is difficult enough to provide correlations between processes and outcomes, much less to ascribe self-reports as causal. Too simple, an evaluation increases the margin of error when interpreting the findings. So what are the basic takeaways from this discussion on evaluation?

Create an infrastructure and culture of evaluation and assessment within the library. This may require strategic and operational reviews of the library, its institution, and accrediting bodies. It may require time and money for staff development and training so staff have the skills and expertise to conduct the types of evaluation that meet requirements for achieving educational/learning outcomes across user groups.

Create evaluations with realistic outcomes that address changes in user behaviors that increase their knowledge, skills, and abilities. As noted earlier, counting how many students ‘like’ a class does not mean they have gained the requisite knowledge, skills, or abilities. Measuring how student confidence in creating effective search strategies has led to the acquisition of requisite skills to meet specified learning outcomes is better.

Clarify definitions and provide consistency in the language of evaluation within our libraries. Whether it is an evaluation framework developed by the Institute for Museum and Library Services, the Medical Library...
Association, the Association of Research Libraries, or the larger institution (such as a university), establish the authoritative resources for use by the library and its staff.

Each time a new framework is adopted, the scope of work should state that the framework has changed, create a crosswalk for any definitional changes and identify implications for its use for evaluation of a current program assessed under the old framework. Such a crosswalk can tie together longitudinal evaluations of programs, an ongoing issue when long-term programs (or projects) suddenly are reevaluated under new criteria without any reference to the original intent of the program. An example is programs developed under the 2000 ACRL Information Literacy Competency Standards for Higher Education versus reevaluation for the 2016 ACRL Framework for Information Literacy for Higher Education. Hence, the context (how it will be used), content (necessary knowledge), and cognitive/affective processes (such as adaptive reasoning, which link content and context, and procedural fluency) that are essential to an understanding of literacy as it is framed today translate or build upon necessary programmatic changes (Hanson and Levin 2013).

In closing, we are not saying librarians shouldn’t be counting. What we are saying is that librarians should count as carefully and contextually as we can to ensure we conduct program evaluations that have sound theoretical underpinnings, realistic claims, solid process, and well-substantiated inferences. Methodologies and data collection will continue to develop; however, using contextualized, thorough, and practical approaches add analytic value and allow us to compare skills acquisition with skills acquisitions, and correlate processes and outcomes.

### Bibliography


Lancaster, F. Willfrid. *If You Want to Evaluate Your Library ...* Champaign, IL: University of Illinois, Graduate School of Library and Information Science, 1988.


