

Preprint of: Nicholson, S. (2004). A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of Documentation* 60(2).

A Conceptual Framework for the Holistic Measurement and Cumulative Evaluation of Library Services

Scott Nicholson, School of Information Studies, Syracuse University
4-127 Center for Science and Technology, Syracuse, NY 13244
scott@scottnicholson.com 315-443-1640

Keywords:

Libraries, Evaluation, Measurement, Library Frameworks, Library science, Systems theory, User studies

Abstract:

This conceptual piece presents a framework to aid libraries in gaining a more thorough and holistic understanding of their users and services. Through a presentation of the history of library evaluation, a measurement matrix is developed that demonstrates the relationship between the topics and perspectives of measurement. These measurements are then combined through evaluation criteria, and then different participants in the library system view those criteria for decision-making. By implementing this framework for holistic measurement and cumulative evaluation, library evaluators can gain a more holistic knowledge of the library system and library administrators can be better informed for their decision-making processes.

“The characteristic way of management that we have taught in the Western world is [to] take a complex system, divide it into parts and then try to manage each part as well as possible. And if that’s done, the system as a whole will behave well. That’s absolutely

false, because it's possible to improve the performance of each part taken separately and destroy the system at the same time." (Ackoff, 1993)

In order to make informed decisions and justify services, librarians should evaluate their offerings on a regular basis (Bawden, 1990). In reality, many evaluations occur because of a problem or report requiring immediate management involvement. These last-minute evaluations are akin to modern emergency-room medicine: just as many patients wait until the symptoms become unbearable before seeking treatment, many library decision-makers wait until problems force a rapid evaluation. Just as the goal of holistic medicine is reaching a state of wellness for the entire body, the goal of holistic evaluation is reaching a state of wellness for the entire library. While the subsystems of a human body are more closely entwined than the subsystems of a library, enough connections exist between the library subsystems to give this comparison validity. One of the fundamental components to holistic theory is that individual components can be combined to produce something beyond the sum of those components (Wilbur, n.d.). In the context of measurement and evaluation, it means that a more thorough knowledge and understanding of a system can be gained from combining different measures than can be derived from taking those measures separately. This conceptual framework for holistic evaluation presents a matrix of perspectives for library measurement and evaluation developed from prior research. In applying this framework, decision-makers not only better understand their system but also can respond to problems more quickly from a preexisting environment of evaluation.

One of the challenges that decision-makers face is the broad array of contexts under which this measurement and evaluation can take place. It is more convenient when evaluating library services to wait for a problem, as the problem will have an apparent context for the evaluation. According to systems theory (Bertalanffy, 1962), however, the complexity of users, systems, and processes is all connected; problems in one area may come from another area of the system. A one-time evaluation of a single portion of the system, while convenient, will leave out portions of the system that could lend other crucial viewpoints for decision-making. Therefore, this holistic framework will guide evaluators to the

consideration of the entire system and not just the problem areas. Goodall discussed a significant gap in library evaluation and supports the need system-wide evaluation: “the total library system is rarely considered in evaluation and ... there is often no clear understanding of the purpose of an evaluation” (Goodall, 1988, p. 129).

Traditional evaluations, which focus on a single user group or system, may not provide managers of these services with the information needed to make effective evidence-based decisions; in fact, one of the implications of systems theory is that the entire system must be evaluated in order to fully understand effects of changes in one component of the system (Ackoff, 1993). Since evaluation requires a viewpoint, it is important to ensure that viewpoints of various participant roles are represented in the decision-making process. In order to aid this important process, this research presents a framework to guide holistic measurement and cumulative evaluation of library services.

Key Definitions and Selected Prior Work

Two essential definitions for this work are measurement and evaluation. The metadiscipline of evaluation has developed as a field in the last few decades (Scriven, 1991); this has resulted in a specific distinction between these two terms. These terms are distinct and have been well-defined. Measurement is the “determination of the magnitude of a quantity” (Scriven, 1991, p. 226) while evaluation is “the process of determining the merit, worth, or value of something, or the product of that process” (Scriven, 1991, p. 139). A simpler definition, but no easier to accomplish, is that “Evaluation consists of comparing ‘what is’ to ‘what ought to be’” (p. 3, Van House, Weil, & McClure 1990). In the framework presented in this paper, these terms are used judiciously. Measurement alone will not aid in the feedback loop of a system; instead, measurement is just a precursor for evaluation in order to fully understand a system. Although the title of this work contains the term “evaluation,” the framework will actually start with evaluation’s precursor, measurement.

Evaluation, as seen here, is supported by social realist evaluation theory. This theory states that “outcomes (i.e. outcome patterns or regularities. . .) follow from mechanisms (sets of internally-related practices and/or objects) acting in contingently configured contexts” (Spasser, n.d., p. 12). These evaluations focus on the reality of the systems, but examine that reality in different contexts. The purpose of that evaluation is to understand the changes brought about by the introduction of the system; therefore, the goal is to not just create the library but to understand what the system allows the user to do and how it impacts those users. Spasser (n.d.), explores social realist evaluation as applied to digital libraries and concludes that “the strength of evaluation research depends upon the perspicacity of its view of explanation” (p. 13).

A few other scholars have presented library evaluation frameworks that focus on multiple perspectives. Saracevic (2000) discussed the evaluation of digital libraries and presented a set of elements for evaluation. This list consisted of the different aspects of a digital library, including traditional library elements such as collections, access, preservation, and use; elements from computer systems such as networks and security; and elements from the management of services such as integration, cooperation, staffing, and costs. He goes on to present the context of evaluations as being user-centered (at the society, institution, individual, and interface level) or system-centered (at the engineering, processing, or content level). This concept of user and system perspectives forms one of the basic dimensions of the measurement component of this framework.

Hernon and McClure(1990) presented four different levels of analysis for evaluation: individual, program, organizational, and societal. They argue that it is important for a library to evaluate its performance at the organizational and programmatic level. All four perspectives are important to consider; in fact, these four perspectives are the basis for the evaluative component of the framework developed here.

Cronin (1982) presented an evaluation matrix for library services with groups across the top (user, management, and sponsor) and types of evaluation (cost,

effectiveness, and benefits) down the side. Griffiths and King(1993) also use a two-dimensional framework, this one targeted toward special libraries. Their dimensions are the object of measurement: entire library, functions performed, services and products, activities, and resources, and the evaluation perspectives: library, user, organization, industry, sector/society. They then map the measures of input, output, usage, and outcomes onto the perspectives to determine the measures to be used. These two frameworks are expanded and combined into the multi-dimensional framework for holistic evaluation developed here.

Baker and Lancaster (1991) have written a thorough and practical text on this topic. In *The Measurement and Evaluation of Library Services*, the authors present a number of different classifications of types of evaluation. They present both microevaluation/ macroevaluation and formative/summative evaluation as ways of thinking about types of evaluation. The strongest theme in the work, however, is based on Lancaster's famous cost-performance-benefit studies, where effectiveness, cost-effectiveness, and cost-benefit aspects of evaluation are considered. The holistic evaluation framework developed here lays the groundwork for all of these evaluation types; after a librarian selects the appropriate perspectives from which to examine library system, Baker and Lancaster's work is recommended as a resource for selecting the particular types of evaluation criteria to use in that situation.

Development of a Library Measurement and Evaluation Framework

“The aim of theorizing is to unify and systemize knowledge” (Kaplan, 2002, p. 310).

This framework will be built in two parts. First, a matrix of topics and perspectives for measurement will be built, using examples of library measurement and evaluation studies as appropriate. Much of the inspiration for measurement comes from past evaluation studies, as evaluation must start with measurement. Therefore, when presenting evaluative examples from the past in

developing this matrix for measurement topics and perspectives, the measurement portion from past evaluation studies is considered. Understanding the topic and perspective of measurement clearly before evaluating is essential in developing a holistic understanding of a library.

In the second part of this framework, that matrix will be extended to introduce evaluation viewpoints in order to encourage decision-makers to take multiple viewpoints into account. The final framework will guide library evaluators and decisions-makers in considering multiple perspectives and viewpoints to gain a holistic knowledge of the library system.

Building the Base: Developing a Matrix of Measurement Topics and Perspectives

Traditional Library Measurement and Evaluation

Traditional forms of library evaluation do not involve users directly and are therefore internal. Early forms of library evaluation started with measurements based on library staff, processes, or systems and not the user (Dervin & Nilan, 1986). These tools were employed to improve library procedures and make the library more efficient. This type of evaluation is still important, as a library that does not function effectively and efficiently will not be able to succeed; however, these measures alone are not sufficient.

Another form of traditional library evaluation is that based on the measurement of the success of an information retrieval system or service. The Cranfield studies, best known for the development of precision and recall measures, did not involve user evaluations; instead, the “relevance” decisions were made by researchers (Swanson, 1965). Hernon and McClure (1986), in their well-known evaluation of reference services, used library researchers to simulate the user experience. These methods may provide a convenient way to quickly judge the success of a system and can inspire future studies, but are all based on an internal view of the library system.

Librarians still use measurement from the perspective of the system today. Those

in collection development measure the coverage of libraries through the use of book lists, reference librarians count the frequency of question types to measure their services, and designers of digital libraries work to meet pre-established standards. These lists and categories may come from library staff or be representative of user needs, but do not reflect any particular user. In this framework, these types of measurements are seen as originating from the internal perspective of the library system.

This topic of library system can include aspects of the collection, the organizational scheme, and computer interfaces; in addition, it might include the library staff and facilities. Therefore, the concept of “library system” goes far beyond a computer system. Instead, the “library system” refers to everything that is part of the offerings of the library.

Therefore, the first quadrant of the measurement matrix is the *Internal Perspective of the Library System*.

	Topic
Perspective <i>Internal (Library)</i>	<i>Library System</i>
	Procedures Standards

Table 1: Measurement Matrix, Quadrant 1: Internal Perspective of the Library System.

Importance of User-based Measurement

In 1968, Orr, Pings, Pizer, and Olson presented a set of tools to aid in the management of library services. Part of their conceptual framework was that the user was at the center of the evaluation. Users saw a system as a “black box,” and evaluation from their perspective did not involve the types of process-related evaluation done in the past. Instead, the user’s perspective – an external perspective - of the system should drive the evaluation.

Taylor introduced the importance of users in information systems with a three-part model, that consisted of a system that adds value to the items in the system, a

set of users who can judge the quality of the output, and an interface where the two meet (1986). He presents the importance of the user in this system, as they drive the measures of quality and the value-added components of the system.

One of the common ways of using judgments from a user is through the inclusion of relevance decisions. Over time, the definition of relevance has been explored by many scholars. Schamber, in her 1994 ARIST chapter, summarized these debates through a three-tiered view of relevance. The system view of relevance is based upon the successful match between the terms in the query and the terms in the documents in the system. The information view of relevance introduces the concept of aboutness, which is based on a content match between the query and the documents. Finally, the situation view of relevance takes into account the individual user and the situation in which the information is to be used (Schamber, 1994). Measurements using the system view and the information view of relevance belong in Quadrant 1 of this matrix, as both come from an internal view; however, the situation view of relevance aids in the introduction of this external, user-based, perspective.

Another type of user-based evaluation of a system is a usability study. The discipline of human-computer interaction (or computer-human interaction, depending upon who is doing the research) focuses on how a user works with a computer system and is essential to those working with digital library services. Not all usability studies are computer-based, however; libraries that exist in a physical space have must be evaluated by users to understand the effectiveness of the layout of space, organization of physical resources, and usefulness of directional resources.

Library evaluators turn to users in a number of ways. The LIBQUAL study focuses on the expectations of users. Designers of digital library services and other computer-based interfaces employ usability studies to understand how individuals work with the online representations of the library. Focus groups are a common way for evaluators to get information directly from a user group. Introducing the external viewpoints of users allow evaluators to look beyond the

constraints of a system and understand what individuals need from a library service.

In order to include the input from users about the system, the measurement matrix is extended to include a new perspective: the *External Perspective of the Library System*.

Topic	
Perspective <i>Internal (Library)</i>	<i>Library System</i>
	Procedures Standards
<i>External (User)</i>	Aboutness Usability

Table 2: Measurement Matrix, Quadrant 2: External Perspective of the Library System.

Expanding the Viewpoints of Users

One cross-disciplinary concept about the perspective of subjects in an evaluation is the difference between emic and etic measures. Originating in linguistics, this concept began as a way to distinguish between differences in language; emic differences are those differences in meaning coming from the speaker, while etic differences are those perceived by the external researcher (Pike, 1982). This concept of emic and etic measures has been adapted for use in other fields, and now inspires this new dimension; the etic measures are those perceived by the library and the researchers, while the emic measures can only come from the user of the system.

Dervin and Nilan presented a summary of research about the importance of including the perspective of users in library evaluation. They presented two paradigms: the first is the traditional paradigm, where “information is seen as objective and users are seen as input-output processors of information”(p. 17, 1986) and that evaluation from this perspective focuses only on the “externally observable dimensions of behavior and events” (p. 17, 1986). The alternative

paradigm is to bring the user into the evaluation and involve their viewpoint, based upon the concept that different users will make sense of an information situation in different ways. This paradigm focuses on “what leads up to and what follows intersections with systems” (p. 17, 1986).

One traditional method of evaluating library services is to examine the concepts of quality and value. Orr presented two basic questions to understand this division: “‘How good is the service?’ and ‘How much good does it do?’” (p. 317, 1973). In order to perform this type of evaluation, librarians must take a different type of measurement from users. Instead of focusing only on the performance of the system, librarians must also consider the user’s viewpoint of their use experience.

Several scholars have looked at ways to measure the user’s view of their use. Dervin’s Sense-Making methodology seeks to measure the user’s information situation, gaps in knowledge that prevent the user from continuing, and how the library aided in resolving that gap (1992). Kulthau (1991) examined the phases that a user passes through during the information seeking process through user interviews and used this to measure the user’s informational and emotional state. Belkin and Kwasnik talked with users about their Anomalous States of Knowledge (ASKs) and measured aspects of the resulting search and usefulness of results to evaluate an information retrieval system (1986). Measuring the user’s view of their use of the system allows for a more holistic understanding of user needs.

Information retrieved from a library may have value to a user even if it does not match the query presented by the user. In a print library, browsing the shelves nearby may allow a user to find works that are useful but not relevant to the original query. The same effect happens on the Internet; a user might be searching for information on one topic and see a resource or link in that digital library that answers a different information need. Kwasnik (1992) looked at the way users browse through information spaces by identifying a set of key activities that occur during browsing. Measuring the browsing activity can be essential to a

library system seeking to show use of a system beyond the traditional measures of circulation (or downloading, in a digital library).

Erdelez (1997) uses the term “information encountering” for information-seeking behavior that goes to topics other than the initial information need. The idea behind information encountering is that library users have a number of information-seeking needs at any one time. There are certain types of people, known as “encounterers”, who will notice information discovered serendipitously that meet another need they happen to have. Therefore, in order to show more benefits to users of a library system, librarians must employ measures that elicit the amount of useful information gained during a session, regardless of the initial information need. To understand the full impact of the information provided by the library, users may need to be contacted some time after their session with the library.

All of these theories point to the need to split the external measurement into two categories: measurement based on the user’s view of the system and measurement based on the user’s view of the use experience. The “use experience” may go beyond the time spent interacting with that library; measuring this may require working with the user before they start their interactions with the library and following up with users well after their library interactions. These post-transactional measurements are crucial to understanding the larger picture of how the library services are being used.

Most of these measures can only be captured through direct contact with the user; it is impossible to accurately ascertain which works a user found valuable without asking that user. One method of detecting scholarly value of a work indirectly is through bibliometric analysis. By examining the user’s written citation/linking behavior (either in print or online) after they interact with the library, evaluators can show the impact of their library. Methods from the disciplines of bibliometrics, informetrics, and scientometrics can be used to document the value of resources provided by a library (Vrana, 2002).

In order to take this new topic of measurement into account, the measurement matrix will be extended to include a new quadrant in a new column: the *External Perspective of Use*.

Perspective	Topic	
	<i>Library System</i>	<i>Use</i>
<i>Internal (Library System)</i>	Processes Standards	
<i>External (User)</i>	Aboutness Usability	Knowledge states Citations to materials

Table 3: Measurement Matrix, Quadrant 3: External Perspective of Use.

Including the Artifacts of Use

The final quadrant of this framework is based on the internal view of the use of the library. This type of evaluation explores the interaction that the user had with the system. It does not discover what the patron wanted to do, nor what they could have done; however, it does capture what the user actually did (as compared to what they say that they did). Understanding this piece bridges the gap between “what the system contains” and “what the user reported about a search”; the internal view of use tells evaluators “how did the user manipulate the current system”.

One traditional form of this measurement is tracking physical access to materials: patrons are asked to not reshelve materials that are used; the evaluators would look for patterns in these measurements to aid in purchasing decisions. Another traditional method of this type of measurement is to look at patterns in circulation; this is much easier now that library management systems track circulation rather than requiring evaluators to examine the check-out cards for use information.

As computer systems became integrated into the library, data-based artifacts of system use were available. Typical measures of these artifacts involve basic frequencies: common reports include amount of circulated materials (and

uncirculated materials), demographic information of users, and number of online catalog and database searches. While these reports were available, and some times even produced regularly in long standardized reports, they were not always used in library decision-making.

Griffiths and King found this to be true in special libraries, and came to the realization that while libraries may use appropriate measures involving budgets and collections, they did not have much information on their user base. For example, many of these libraries did not know the size of their user base or the populations they were to serve. The researchers developed measures to aid in the analysis of special libraries (1993).

As OPACs (online public access catalogs) became more popular, so did methods of measuring and evaluating the use of those catalogs through the transaction logs left behind. Transaction Log Analysis (TLA) became a popular area of research in the 1980's; Peters(1993) provides a detailed literature review on TLA. These concepts of measuring aspects of a user's behavior in a library system continue today through Web log analysis. Libraries providing Web-based services can track similar paths of use through analysis of Web logs. Many of these studies look for common paths and typical searches, as well as exploring the searches performed.

One significant advantage of this type of data collection over traditional user studies is the large quantity of data. Data captured from traditional library systems may only contain partial information (e.g. circulation records do not represent all use of a print work); however, in digital library services, evaluators have the ability to track everything the user does within the constraints of the system. By collecting this data over time and matching demographic information with transactional data, the bibliomining process, or data warehousing and data mining for libraries, can be used to then discover patterns of use (Nicholson and Stanton, 2003).

These bibliomining studies do not replace other types of user studies; rather, they

enhance them by providing a more complete view of the user's experience with the library. Large-scale studies such as this also can allow librarians to see overall trends that may not be evident from smaller samples. A library manager that judges success of services from complaints and complements sees only the extremes; without understanding the overall trends, it is difficult, if not impossible, for a manager to put these comments in the appropriate perspective.

One of the current grand challenges for the future of librarianship according to Buckland (2003) is to gain a better understanding of different communities of library users. One difficulty with this challenge is that there are many variables upon which to group library users into communities. Bibliomining is designed to extract meaningful patterns of use from large amounts of transactional data; this data-driven method allows libraries to determine the key variables that cause true differentiation between communities. Understanding these differences allows libraries to offer and personalize services to meet the needs of more communities.

Kaplan summarizes this view succinctly when discussing the models of cybernetics and information theory by stating that "the purposiveness of behavior can be simulated by artifacts" (2002, p. 292). By examining these data-based artifacts of behavior, evaluators can understand patterns of use. Therefore, the final quadrant of the measurement matrix is the *Internal Perspective of Use*.

Perspective	Topic	
	<i>Library System</i>	<i>Use</i>
<i>Internal (Library System)</i>	Procedures Standards	Recorded interactions with interface & materials
<i>External (User)</i>	Aboutness Usability	Knowledge states User citations to materials

Table 4: Complete Measurement Matrix

Summary of Measurement Matrix

The purpose of this matrix is to aid library evaluators in choosing targets for measurement that will help in the understanding of the library system from a

more holistic view. There are four parts to the complete matrix:

- the internal view of the library system (*what does the library system consist of?*), which does not involve users and compares components of the library system to some type of standard;
- the external view of the library system (*how effective is the library system?*), where the user presents a query to the library and examines the usability of the system and the aboutness of the results presented by the library;
- the external view of use (*how useful is the library system?*), where the user presents the overall usefulness of information gained through the library, either through elicitation by an evaluator or by citing/linking to library works
- the internal view of use (*how is the library system manipulated?*), where the data-based behavioral artifacts of interactions between users and a system are analyzed to understand how a system is manipulated

The most important implication of this framework is that there is something to learn from each perspective and topic. There is not a single measure that can be taken that represents the library; multiple measures are needed to holistically understand the entire library system. An important note is that the same measurement situation can address multiple quadrants with different measures: for example, a user could use the system, judge the items for aboutness, discuss the usefulness of the items, and the librarian can examine how the user worked with the system. *Using a variety of measurements for evaluation ensures a more holistic understanding of that library.*

Methods for Measurement

Once the topics have been selected, the researcher needs to choose methodologies that are appropriate for that topic. It is outside the scope of the present work to present the suite of methodologies available for library measurement; researchers such as Kania (1989), Baker and Lancaster (1991), McClure (1994), Cullen (1998), Bertot, McClure, and Ryan (2000), and Tenopir (2003) have provided works that discuss different ways of collecting these measures. One of the findings of Tenopir (2003) was that many library studies attempt to draw

conclusions that are not possible through the measures used to collect data. For example, data from search logs cannot tell a researcher which documents a user found to be relevant; only the user can report relevance of documents to their information need. Table 6 contains some of the methodologies that can be used (with appropriate questions) to collect measures from each quadrant of the Measurement Matrix.

Perspective	Topic	
	<i>Library System</i>	<i>Use</i>
<i>Internal (Library System)</i>	<ul style="list-style-type: none"> • Staff surveys and interviews • Audits of collections, systems, or staff 	<ul style="list-style-type: none"> • Bibliomining • Transaction / Web Log Analysis • Observation of user behavior
<i>External (User)</i>	<ul style="list-style-type: none"> • Surveys and interviews • Talk-alouds and in-process feedback mechanisms • Focus groups 	<ul style="list-style-type: none"> • Surveys and interviews • Focus Groups • User citation tracking

Table 5: Methodologies appropriate for each quadrant of the Measurement Matrix.

Moving from Measurement to Evaluation

After measures have been collected, these measures must be evaluated. Measuring without evaluating is a common problem with automated reporting tools. Measurement produces data; however, evaluation creates information. The evaluation involves some method of judgment about the collected measures and metrics through some criteria. Judgment requires a viewpoint; different viewpoints may lead to different results. Therefore, selection of multiple evaluation criteria and the viewpoints is critical in gaining a more thorough understanding of the library.

This problem can be exemplified by different definitions for quality presented by Brophy and Couling. Some definitions for quality come from the matching of a

system to specified tolerances; while other measures of quality are based upon the needs of the user. In addition, they present different definitions of effectiveness based upon different stakeholders, leading to their “Coalition Approach to Effectiveness” (1996, p. 142). The problem causing the need for these many definitions is a lack of a conceptual framework for the context of the evaluations. Thus, there are two parts to an evaluation: the evaluation criteria and the evaluation viewpoint.

Selecting Evaluation Criteria

Numerous researchers have presented different frameworks of evaluation criteria. Lancaster presented one of the most commonly accepted frameworks for evaluation consisting of three tiers: effectiveness, cost-effectiveness, and cost-benefit. Effectiveness is “how well the system is satisfying its objectives” (1971/1978, p. 23). Once effectiveness is measured, the cost of the service can be introduced to examine the cost-effectiveness of the service. Finally, this framework recognizes that effectiveness and benefits are not the same; therefore, cost-benefit is evaluating a service based upon the cost as compared to the benefits provided by that service (Lancaster, 1971/1978).

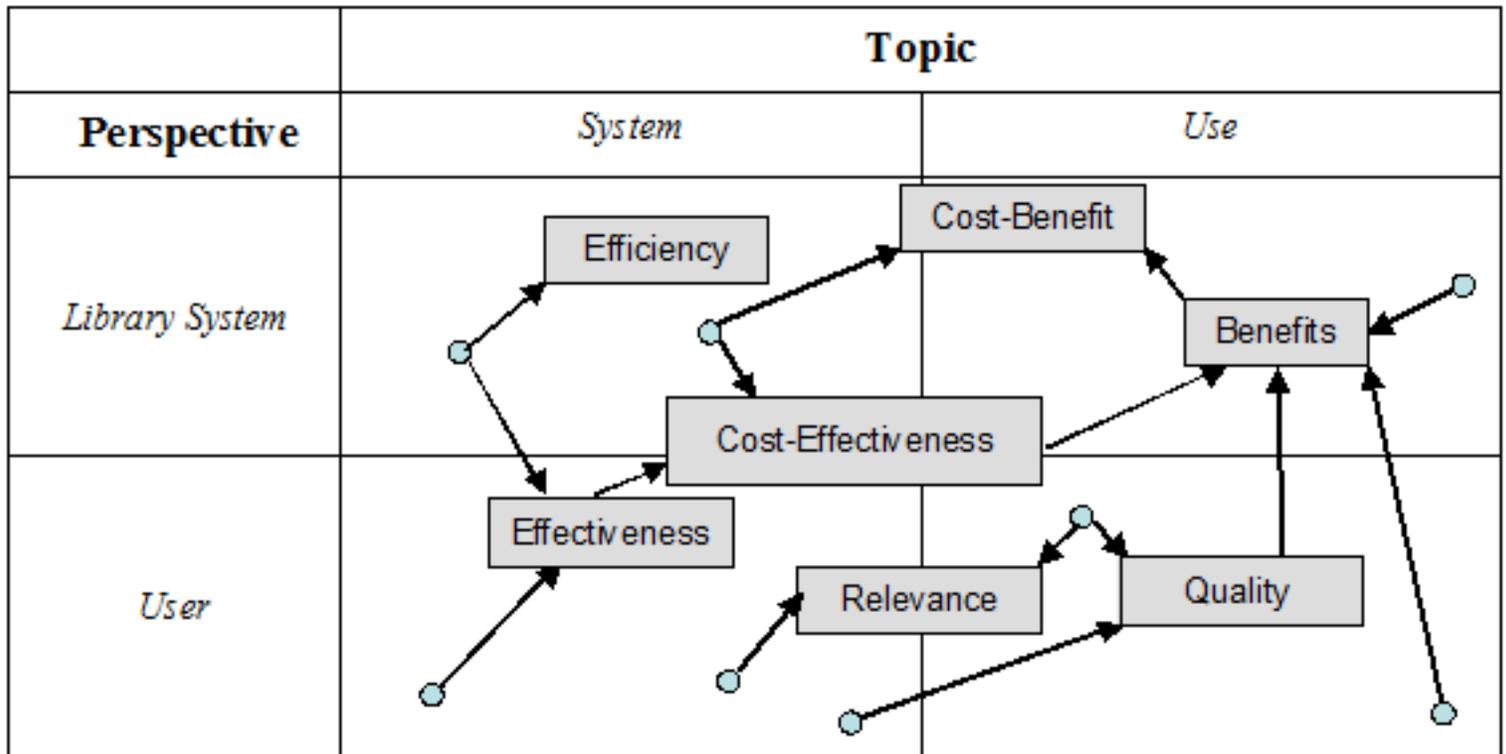


Figure 1: Mapping Example Evaluation Criteria to the Measurement Matrix

Each of these criteria can be determined from different measures. Effectiveness, for example, combines a measure from the internal view of the system (objectives) to another measure (most likely one from the external view if that objective is based upon the user). Introducing the cost of the service brings in another measurement point from the internal view of the system. The benefits will come from yet another measurement point from one of the quadrants related to users. Another common criterion, quality, requires the user's view of not only the library system but also the overall use experience. Therefore, different evaluation criteria can pull from different quadrants of measurement, making the holistic understanding of the library easier to accomplish (see Figure 1 for an example of mapping criteria to the matrix).

The same process can be used with any other evaluation criteria by mapping criteria to one or more measures. Libraries seeking to implement this model can start here, by taking evaluation criteria currently employed and mapping them to the matrix. This will guide library evaluators in deciding what other types of

measurements and evaluative criteria are needed for the holistic measurement of their library. Once the library begins to implement evaluations that require collection of different measurement types, the data resources will be available to rapidly extend the types of measures and evaluations available for library justification and decision-making. Using this holistic framework as the central planning tool for library measurement and evaluation makes allows the relationship between measurements and evaluation criteria to flourish and encourages a more holistic understanding of library services.

Considering Different Evaluation Viewpoints

The same evaluation criteria will be judged in different ways by different participants in the process. In order to gain a holistic understanding of the evaluation, the viewpoints from different groups must be taken into perspective. For example, the criterion of cost-benefit may be judged differently by a user, a librarian, and the funding agency for the library. Therefore, it is important to be aware of the viewpoint of the group doing the evaluation and ensure different groups who might be affected through decisions made from the evaluation can participate in the process (Brophy and Couling, 1996).

In his classic 1997 JASIS article, Saracevic presented a very important distinction surrounding user-based relevance of an IR system. There are two important, but distinct, measures for search results from the user's perspective. He presented relevance as the match between a query and a document, and then went beyond that to *value*, which is the usefulness of the material. He stated that a document must be relevant to a user to have value, but relevance by itself is not enough to merit a valuable match (Saracevic, 1997). This concept embodies the idea of the split between measurement and the evaluation and the importance of the user in this process.

Many scholars argue that the user perspective is the most important aspect in library evaluation (Dervin & Nilan, 1996). After all, if the library services are not for the users, then who are they for? Because of this, the first evaluation

viewpoint that should be taken into account is the user evaluation. For each measure, the user's input, both formally through library friends groups and focus groups and informally through conversations with the users should be collected first. Therefore, the user can evaluate the measures collected through the evaluation criteria. Other participants in the process will then have not only the data from the measurement and evaluation criteria, but also the user's viewpoint on that evaluation.

Therefore, library users are represented twice in this process. First, the users participate by producing some of the measurements gathered, i.e. the bottom row of the measurement matrix. Then, users are tapped again in the evaluation process. These might be the same users or different users, but now instead of working with the library services, the users are asked to evaluate the library services through the lens of collected measures and evaluation frameworks.

The next level of evaluation are the library personnel, who will take into account the data collected as well as the user's evaluation of that data. Within the library staff, there may be divisions based upon the context of the library; a common division in a physical library would be library technicians, librarians, and library managers. Each of these groups may evaluate a measure in a different way and therefore, it is important to have cross-library representation in this process

The third level is made up of the decision-makers: funding agencies, administration, and policy makers (as appropriate). Most libraries report to a higher body for their funding and/or administration. It is important that these decision-makers know the not only the data collected and the criteria for evaluation, but also the evaluation of that measure by both users and librarians. Employing this framework for holistic measurement and cumulative evaluation will allow the systematic collection of measurement topics and viewpoints as well as evaluative perspectives in order to ensure that decision-makers have a thorough understanding of the library system.

Therefore, in order to complete this framework, the evaluation viewpoints are

presented in a rising pyramid (See Figure 2). A pyramid was selected for this model for several reasons. The base of the pyramid is the measurement matrix, as those measurements are the base of this entire process. The evaluation criteria are used to combine and collect the measurements to make them easier to understand. Moving up the pyramid, these criteria are evaluated by users, library personnel and then the policy makers, administration, and funding agencies. As the levels get higher, the number of people involved in the process shrinks; this is represented by the shrinking levels in the pyramid. However, the impact by individuals on the library system increases as the number of people represented by each level decreases. In addition, the evaluations performed at each level take into account everything below it. After decisions are made, they are applied by the library personnel, affect the users and are eventually cause for measurement and evaluation, then the process starts again (see Figure 3).

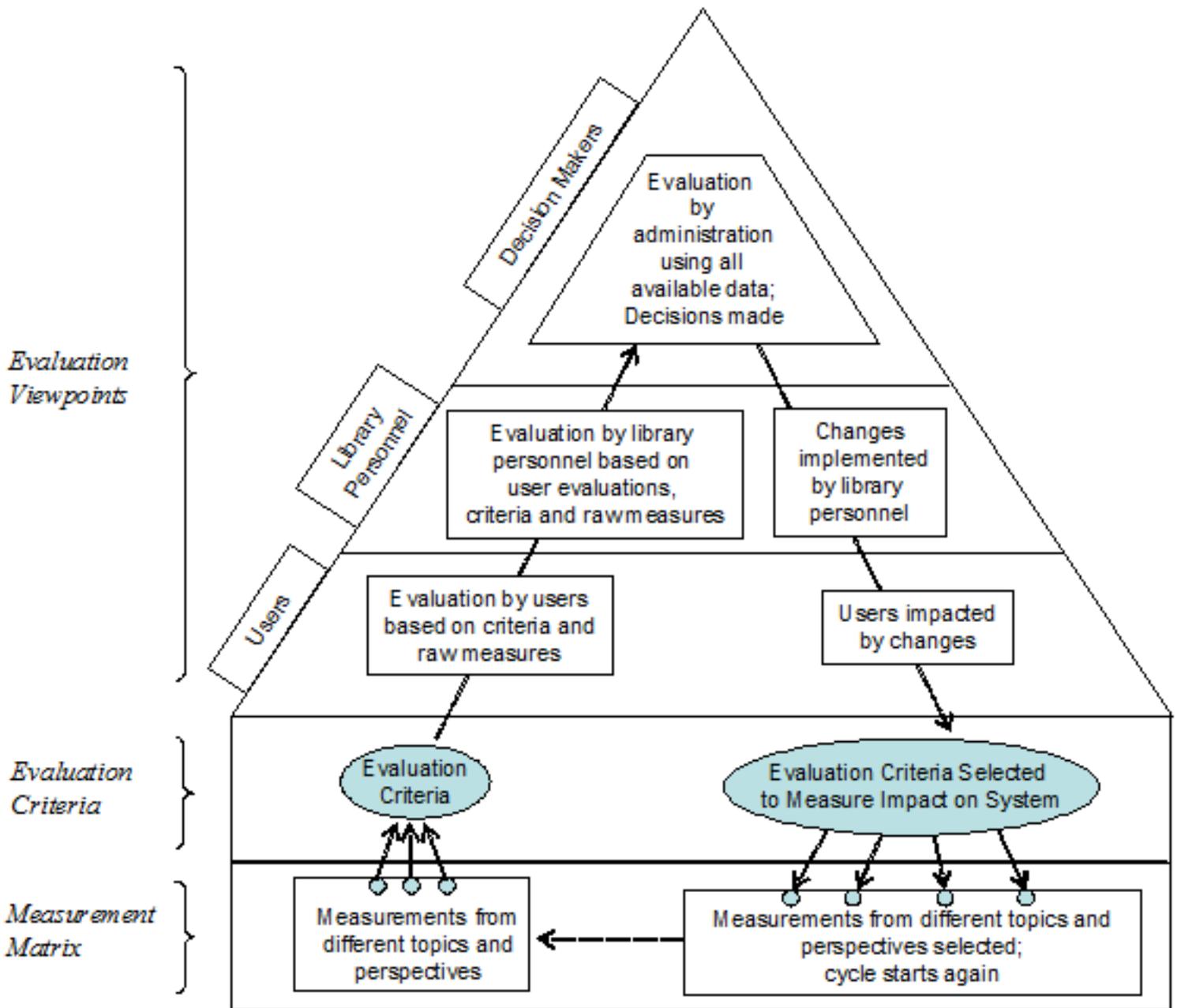


Figure 2: Cycle of Holistic Measurement and Cumulative Evaluation of Library Services

One interesting note about the division between “measurement” and “evaluation” changes between evaluation viewpoints in the pyramid. The user evaluates the measurements collected with the help of evaluation criteria. The library personnel perform their evaluations using not only the evaluation criteria on the measurements collected but also the user evaluations. To the library personnel, the user evaluations are measures to be collected without adding value. Finally, the decision-makers make their evaluations based upon the measures of collected

data, user evaluations, and library personnel evaluations. The definitions of “measurement” and “evaluation” change based upon the viewpoint and place in the cycle of the person doing the evaluating.

Implementation of Measurement Matrix and Evaluation Framework

For a library new to these concepts, this framework may seem overwhelmingly time-consuming and expensive to implement. However, this framework should be seen as a guide to the decision-making process about measurement and evaluation rather than the first step in the process. The first step is to examine what types of measurement and evaluation are currently done at the library and map that onto the framework structures. This will provide librarians with an idea of what they are currently measuring, how those measurements relate to each other, the role that evaluation criteria is currently playing in the perception of the measurement, and who is examining the evaluation criteria.

Once this audit and mapping of measurement and evaluation components is complete, gaps will become apparent. These gaps are then used to inform decisions about what studies to select and implement in order to improve understanding of the library. Each new addition to these evaluation procedures moves the library one step closer to a holistic understanding of their services. Therefore, over time, the library will build a set of tools and procedures that represent the multiple topics and perspectives of the measurement matrix and viewpoints of the evaluation pyramid.

Conclusion

In order to gain a thorough understanding of a library system, decision-makers must work to gather measures from different areas of the library system using different perspectives. These measures should then be evaluated by users and library personnel; all of this data can then be taken into account before making decisions, changing policies, or issuing funding for library services. Bypassing these processes guarantees that part of the library system will be neglected in this process, and the resulting decisions can break that interconnected system.

The competing forces and multi-dimensional context surrounding library services, combined with the emergency-based nature of most evaluation projects, can put library evaluators and decision-makers in a situation where they apply convenient evaluation tools to accessible data instead of working to holistically understand the library system. All components of the library function as a single system, and making changes based upon an evaluation of a small component of that system can be problematic.

What is the purpose of this framework? Many library evaluations are driven by problems, and not through a systematic analysis of the situation. As Kaplan said, “a theory is a way of making sense of a disturbing situation so as to allow us most effectively to bring to bear our repertoire of habits, and even more important, to modify habits or discard them altogether, replacing them by new ones as the situation demands” (2002, p. 295). Applying the instrumentalist view of theories as discussed by Kaplan, this framework can be used as “tool(s) of inquiry, and of reflective choice in problematic situations” (2002, p. 306). Instead of rushing to with an evaluative patch to repair a problem, library managers can apply this framework to ensure that different perspectives are taken into account to fully understand library services and make sound decisions.

References

Ackoff, R. (1993). *A Theory of a System for Educators and Managers. The Deming Library, Vol. 21*. Silver Springs, MD :; CC-M Productions. Retrieved July 2, 2003 from <http://www.managementwisdom.com/abdrusac.html>

Baker, S. & 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*, Aldershot, England: Gower.

Belkin, N. & Kwasnik, B. (1986). Using structural representations of Anomalous

- States of Knowledge for choosing document retrieval strategies. In F. Rabitti (ed.), *ACM/SIGIR Conference Proceedings. Pisa: Pisa University*. 11-22.
- Bertalanffy, L. von (1962). General system theory: A critical review. *General Systems* 7, 1-20.
- Bertot, J., McClure, C. & Ryan, J. (2000). *Statistics and Performance Measures for Public Library Networked Services*. Chicago: American Library Association.
- Brophy, P. & Coulling, K. (1996). *Quality Management for Information and Library Managers*. London: Aslib Gower.
- Buckland, M. (2003). Five grand challenges for library research. *Library Trends* 51(4). 675-686.
- Cronin, B. (1982). Taking the measure of service. *ASLIB Proceedings* 34(6/7), 273-294.
- Cullen, Rowena. (1998). Measure for measure: A post modern critique of performance measurement in libraries and information services. Victoria University of Wellington New Zealand, School of Communications and Information Management. Available: ERIC: ED434664.
- Dervin, B. (1992). From the mind's eye of the user: The Sense-Making qualitative-quantitative methodology. In J. D. Glazier & R. R. Powell (Eds.), *Qualitative research in information management* (pp. 61-84). Englewood, CO: Libraries Unlimited.
- Dervin, B. & Nilan, M. (1986). Information needs and uses. In Williams, M. E. (Ed.) *Annual Review of Information Science and Technology* (21). Washington, DC: American Society for Information Science, 3-33.
- Erdelez, S. (1997). Information encountering: A conceptual framework for

accidental information discovery. In P. Vakkari, R. Savolainen, & B. Dervin (Eds.), *Information Seeking in Context. Proceedings of An International Conference on Research in Information Needs, Seeking and Use in Different Contexts* (pp. 412-421). London: Taylor Graham.

Goodall, D. (1988). Performance measurement: A historical perspective. *Journal of Librarianship* 20(2), 129-144.

Griffiths, J. and King, D. (1993). *Special Libraries: Increasing the Information Edge*. Washington, DC: Special Libraries Association.

Hernon, P., & McClure, C. (1990). *Evaluation and Library Decision Making*. Norwood, NJ: Ablex Publishing Corporation.

Hernon, P., and McClure, C. R. (1986). Unobtrusive reference testing: the 55 percent rule. *Library Journal*, 111(April 15)37-41.

Kania, Antoinette M. (1988). *Performance measures for academic libraries: A twenty year retrospective*. Available: ERIC: ED293540.

Kaplan, A. (2002). *The Conduct of Inquiry: Methodology for Behavioral Science* (2nd printing, Rev. ed.). New Brunswick, New Jersey: Transaction Publishers.

Kwasnik, B. (1992). A descriptive study of the functional components of browsing. In *Engineering for Human-Computer Interaction*, ed. J. Larson and C. Unger, 191 – 202. Amsterdam: Elsevier Science Publishers.

Kuhlthau, C. (1991). Inside the search process: Information seeking from the users perspective. *Journal of the American Society for Information Science* 42(5). 361-71.

Lancaster, F. W. (1978). The Cost-Effectiveness Analysis of Information Retrieval and Dissemination Systems. In King, D. (Ed.). *Key Papers in the*

Design and Evaluation of Information Systems (pp. 23-38). White Plains, NY: Knowledge Industry Publications, Inc. (Reprinted from *Journal of the American Society for Information Science*, Jan-Feb. 1971, 13-27).

McClure, C. (1994, Spring). User-based data collection techniques and strategies for evaluating networked information services. *Library Trends*, 42 (4), 591-607

Nicholson, S. & Stanton, J. (2003). Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. In Nemati, H. & Barko, C. (Eds.). *Organizational data mining: Leveraging enterprise data resources for optimal performance*. Hershey, PA: Idea Group Publishing.

Peters, T. (1993). The history and development of transaction log analysis. *Library Hi-Tech* 42(11). 41-66.

Pike, K. (1982). *Linguistic Concepts: An Introduction to Tagmemics*. Lincoln: University of Nebraska Press.

Orr, R.H. (1973). Measuring the goodness of library services: A general framework for considering quantitative measures. *Journal of Documentation* 29(3). 315-332.

Orr, R.H., Pings, V., Pizer, I., & Olson, E. (1968). Development of methodology tools for planning and managing library services: I. Project goals and approach. *Bulletin of the Medical Library Association* 56(3). 235-240.

Saracevic, T. (2000). Digital library evaluation: Toward an evolution of concepts. *Library Trends* 49(3), 350-369.

Saracevic, T. & Kantor, P. (1997). Studying the value of library and information services: Part 1, Establishing a theoretical framework. *Journal of the American Society for Information Science*. 48(6), 527-542.

Schamber, L. (1994). Relevance and information behavior. In Williams, M. E. (Ed.), *Annual Review of Information Science and Technology*, 29. Medford, NJ: Learned Information, Inc. 3-48.

Scriven, M. (1991). *Evaluation Thesaurus* (4th edition). Newbury Park: Sage Publications.

Spasser, M. (n.d.). *Realist Activity Theory for Digital Library Evaluation: Conceptual Framework and Case Study*. Retrieved April 22, 2003 from <http://...>

Swanson, D. (1965). The evidence underlying the Cranfield results. *Library Quarterly* 35(1), 1-20.

Taylor, R. (1986). *Value-Added Processes in Information Systems*. Norwood, NJ: Ablex Publishing Corporation.

Tenopir, C. (2003). *Use and users of electronic library resources: An overview and analysis of recent research studies*. Washington, D.C.: Council on Library and Information Resources.

Van House, N., Weil, B., & McClure, C. (1990). *Measuring academic library performance: A practical approach*. CITE PUBLISHER.

Vrana, R. (2002). Digital Libraries - Creating Information Space Excellence: Is It Already Time for Benchmarking?. Paper presented at the 2002 CARNET Users Conference. Retrieved April 20, 2003 from http://public.srce.hr/~rvrana/cuc2002_vrana.pdf

Wilbur, C. (n.d.). *Holistic Method*. Retrieved August 19, 2003 from <http://www.nd.edu/~cwilber/pub/recent/edpehol.html>

Author Biography

Scott Nicholson, PhD, MLIS, is an assistant professor at the School of Information Studies at Syracuse University in Syracuse, New York. His research focuses on the evaluation of library services and the use of bibliomining, or the combination of bibliometrics and data mining, to aid in library decision-making through the development of standards and metrics. He is the founder of the Bibliomining Information Center at bibliomining.org, and can be reached at scott@scottnicholson.com.

Acknowledgement

The author wishes to thank Kenneth Lavender, Barbara Kwasnik, Abby Goodrum, Jeff Pomerantz, and Joanne Silverstein for their support and feedback throughout the creation of this work.