A Framework for Evaluating Digital Library Services

Saveed Choudhury  
Benjamin Hobbs  
Mark Lorie
Johns Hopkins University  
<saveed, bhobbes, mlorie@jhu.edu>

Nicholas Flores  
University of Colorado  
<nicholas.flores@Colorado.edu>

Abstract

This article provides an overview of evaluation studies for libraries, a brief introduction to the CAPM Project, a description of the theoretical background for the CAPM methodology and, finally, a discussion of the implementation of the methodology for the CAPM Project.

Introduction

The level of interest regarding digital libraries has grown steadily as a greater number of institutions, including libraries and museums, consider the possible implications of digital libraries. While there are important, unresolved digital library research and development issues, there is also a concurrent desire to develop strategies for systematic digital library programs based upon results of digital library projects. Digital library programs generally include both digital collections and services that facilitate access, retrieval, and analysis of the collections. This interest reflects growing expectations from patrons and end users. In an ideal world, with unlimited resources, it would be possible to provide a full range of digital library services to all users. In reality, resource constraints require a consideration of priorities. Consequently, it would be useful to evaluate potential benefits, as determined by patrons and end users, regarding digital library services. Even without considering digital library services, Saracevic and Kantor (1997a), and Kyrillidou (1998) have provided compelling reasons for evaluating libraries based on user feedback.

Evaluation of digital library services, based on user feedback, can take many forms depending on the objectives of the analysis. For example, usability testing often focuses on assessing the effectiveness, efficiency and/or satisfaction of the user experience with a particular interface (Nielsen 1993; Norlin and Winters 2002). This type of testing assumes the existence of a prototype or more fully developed version of the interface.

This article focuses on evaluation methodologies that consider the outcomes, impacts, or benefits of library services. The authors have developed an evaluation methodology, based on a multi-attribute, stated-preference economic model, that was utilized to evaluate the Comprehensive Access to Printed Materials (CAPM) project at Johns Hopkins University (JHU). While this "CAPM methodology" was applied to a specific project, the framework
can be used generally for evaluating users' preferences for digital library services. The CAPM methodology provides a framework to prioritize the development of digital library services within an institution, based on users' preferences.

**Evaluation Studies of Libraries**

For a complete evaluation of any system or project, both costs and benefits must be considered. Historically, there has been a greater emphasis on measuring the costs associated with libraries. However, currently there is a growing trend toward examining and emphasizing benefits associated with libraries. The following discussion offers a representative set of such evaluation studies.

The Association of Research Libraries (ARL) has collected statistics related to its member libraries for many years. Previously, these statistics focused primarily on "input" measures such as size of collections or number of staff. Subsequently, ARL considered "output" measures such as circulation statistics. Shim and Kantor (1996) used Data Envelopment Analysis (DEA) to evaluate digital libraries. DEA measures the relative efficiencies of organizations ("decision making units"), given multiple inputs and outputs (Charnes et al., 1978). The measurement of efficiency can apply to a single institution over time, or across multiple institutions. As Shim and Kantor state "an efficient library is defined as the one which produces the same output with less input or, for a given input, produces more output." Shim and Kantor used ARL statistics as the basis for their study. While this approach advanced the notion of evaluation, Kyrillidou (2002) points out that the relationships between inputs and outputs within a library are not necessarily clear. Additionally, Shim and Kantor indicate that libraries must describe how inputs are transformed into services, rather than outputs.

ARL has acknowledged this need through its New Measures Initiative\(^2\), which emphasizes outcomes, impacts, and quality, based on user satisfaction. ARL's E-Metrics project represents an effort to define and collect data on the use and value of electronic resources. ARL's LibQUAL+\(^{TM}\) attempts to measure overall service quality in academic research libraries (Cook et al., 2001). LibQUAL+\(^{TM}\) arose from SERVQUAL, an instrument, based on the gap theory of service quality, which was used to assess private sector institutions. ARL intends to extend LibQUAL+\(^{TM}\) to evaluate digital libraries, through the National Science Foundation's National SMETE Digital Library (NSDL) program.

The aforementioned Kantor, along with Saracevic, (1997a; 1997b) conducted a long-term study to develop a taxonomy of user values for library services and a methodology for applying the taxonomy. They also provide arguments for the importance of user-based evaluations of libraries and assert that users might have difficulty assigning dollar values for library services.

Other examples of user-centric evaluation include Norlin (2000) who evaluated user satisfaction regarding reference services with surveys to gather demographic data, unobtrusive observations of the delivery of reference services, and follow-up focus groups.

Hill et al. (1997) used multiple methods to obtain feedback regarding the Alexandria Digital Library (ADL) at the University of California, Santa Barbara. The study adopted several methods to evaluate user views including: online surveys, ethnographic studies, focus groups, and user comments. The goal of this study was not to compare the value of the system to its costs, but rather to incorporate user feedback in the ongoing design and implementation of the ADL.

Talbot et al. (1998) employed a Likert type survey to evaluate patron satisfaction with various library services at the University of California, San Diego. This survey was conducted in response to a comprehensive change in the library management's philosophy.

Chris Borgman at UCLA has written extensively regarding digital libraries from a
A Framework for Evaluating Digital Library Services

user-centric perspective. One of her recent works, Borgman (2000), provides a multi-disciplinary, holistic, human-centered perspective on the global information infrastructure. Many economists including Hal Varian4, Malcolm Getz5, and Jeff MacKie-Mason6 have examined the evaluation of libraries and information. MacKie-Mason, and others, examined the issue of electronic journals pricing during the Pricing Electronic Access to Knowledge (PEAK) 2000 conference.7

There are a number of studies that adopt multi-attribute, stated-preference techniques, or some variant of them. Crawford (1994) describes a multi-attribute, stated-preference application for evaluating reference services within academic libraries and provides an overview of an earlier study using similar techniques (Halperin and Strazdon 1980). Harless and Allen (1999) utilize contingent valuation methodology (CVM), a subset of multi-attribute, stated-preference techniques, to measure patron benefits of reference desk services. Basically, CVM explores users’ willingness to pay, in dollar values, for varying levels of services. The most widely cited reference for CVM is Mitchell and Carson (1989).

The Harless and Allen paper raises the important distinction between use and option value, concepts that have been developed in the context of environmental goods. Use value reflects the value of benefits as assigned by actual users of specific services. Option value incorporates the additional benefits as determined by users who might use specific services in the future (i.e., individuals who had not used the reference service but still placed a value on its existence). Any evaluation study that focuses only on individuals who use a specific service (e.g., interviewing only patrons as they leave the reference desk) will most probably underestimate the benefit of the service in question.

Outside of the US, the eVALUEd project team has implemented a questionnaire designed to collect data regarding evaluation methodologies in the UK.8 The questionnaire was offered to the heads of Library/Information Services in Higher Education Institutions in the UK. The goal of eVALUEd is to produce a transferable model for e-library evaluation and to provide training and dissemination in e-library evaluation. The results of this effort should provide an interesting comparison to efforts based in the US.

These studies demonstrate an increasing emphasis on both inter and intra-institutional measures, outcomes rather than inputs, a user-centric perspective, adoption of evaluation techniques from various disciplines, and evaluation of digital libraries. The CAPM methodology reflects these trends and offers a complementary framework for evaluation. Consider that a LibQUAL+™ analysis might identify the gaps in digital library services. The CAPM methodology might then provide the means for prioritizing services to address the gaps. Finally, usability testing could offer the means to ensure appropriate design of interfaces for these services. The CAPM methodology provides one of the first multi-attribute, state-preference analyses of a digital library system that will augment an existing library service, a scenario many libraries might encounter as they develop digital library programs. The following section provides an overview of the CAPM project.

**CAPM Project**

The Comprehensive Access to Printed Materials (CAPM) project began in response to a pressing problem facing many libraries, especially academic research libraries (Choudhury 2001; Suthakorn et al. 2002). Given the increase of electronic resources (and associated infrastructure) within libraries, and ongoing print-based acquisitions, most libraries face major space shortages (Wagner 1995). In response to this space constraint, libraries have often built large, off-site shelving facilities to house portions of their physical collections. Johns Hopkins University (JHU) has implemented such a facility, the Moravia Park Shelving Facility. This facility offers high-density shelving with individual books arranged by similar size and contained within open boxes. Patrons may request items from Moravia Park, which are delivered twice daily. While this strategy addresses the space problem, it eliminates the ability to browse *the* materials shelved in these off-site facilities. Patrons
are actually to browse for the materials in Moravia Park, but not the actual contents themselves.

The CAPM Project was initiated to address this lack of browsability. The CAPM system itself will operate as follows: when a patron requests an item that is shelved off-site, a retrieval robot will automatically navigate the stacks to find the desired item and bring it to a scanning system that includes a robotic page-turner. The patron will have complete remote control to enable scanning pages, viewing images, searching full-text and printing pages. Upon completion of the scanning the retrieval robot will return the book to the stacks or separate the book for physical delivery to the main library building. It is important to note that CAPM differs from existing warehouse-automated retrieval systems within libraries (Hansson 1995; Kirsch 1999) since it is an on-demand and batch scanning system for printed materials in remote locations. Ultimately, the CAPM system will enable patrons, even those outside JHU, to browse materials shelved at the off-site facility making on-demand digitization a reality.

With initial funding from the Council on Library and Information Resources (CLIR), the CAPM Project team conducted an initial technical and economic feasibility study. This study provided evidence that a fully robotic CAPM system was possible, but also confirmed that further economic analysis was necessary. The first phase of the CAPM Project, funded by the Mellon Foundation, comprised the development of a prototype retrieval robot and a concurrent evaluation of potential costs and benefits associated with CAPM implementation. The JHU team will continue development of CAPM with a grant from the National Science Foundation's Information Technology Research program. The evaluation from the first phase provided the necessary context to determine whether it would be beneficial, from an economic perspective, to develop the CAPM system. Before describing the specific results of the evaluation, the next section provides an overview of multi-attribute, stated-preference techniques.

**Multi-attribute, Stated-preference Methods**

While the project team evaluated the specific costs and benefits associated with CAPM implementation, they also considered the appropriateness of multi-attribute, stated-preference methods for evaluating digital library services in general.

Multi-attribute, stated-preference methods feature choice experiments to gather data for modeling user preferences. In the choice experiments, often expressed as surveys, subjects **state** which alternatives (services or features) they most prefer; the alternatives are distinguished by their **multi-attributes**. In designing the choice experiments, it is important to develop credible choices such that subjects (users) have appropriate information to make meaningful choices between alternatives. Additionally, the format for making choices must also be meaningful for subjects. Multi-attribute, stated-preference experiments provide data that are then used to estimate the marginal benefit of each attribute. These experiments are based upon the idea that users receive utility from services, and this utility is specified and measured through rational models.

Fundamentally, multi-attribute, stated-preference techniques acknowledge the need for tradeoffs when making decisions regarding resource allocation. The philosophy underlying these techniques is that users can best assess the appropriate tradeoffs to maximize their utility. One of the criticisms of multi-attribute, stated-preference techniques is that they rely upon choices made by users within hypothetical experiments. However, these methodologies offer additional insight and perspective regarding user preferences for decision-makers, who most probably should consider information from various sources. Ultimately, decision-makers have to make decisions. Multi-attribute, stated-preference techniques provide an additional tool for this purpose.

The multi-attribute, stated-preference technique has been used extensively in marketing research to help predict demand for new products. In the past ten years, this approach has been used increasingly for cost-benefit analysis of public projects as well as natural
resource damage assessments. It is important to note that library services are not offered through private markets, a characteristic shared by public projects or natural resources. Adamowicz et al. (1998) provides an overview of the multi-attribute, stated-preference methodology; Kamakura and Russell (1989) describe an application of the methodology for marketing studies while Wisconsin Department of Natural Resource Services (1999) offers an example application for natural resources.

Implementation for CAPM Project

The CAPM project team designed a multi-attribute, stated-preference experiment based on feedback from focus groups with JHU students, faculty and staff, and input from economists, information scientists, and librarians during a workshop. For the CAPM system, the research team considered two formats for the choice experiments. The first format, consistent with the previous discussion, comprised a multi-attribute choice format that featured varying levels of service on the following dimensions:

1. Presence or absence of digital (scanned) images
2. Presence or absence of full text search with digital (scanned) images
3. "Delivery" time that a user must wait to view a requested item from the Moravia Park Shelving Facility
4. Price per semester for using an alternative system

The second possibility for the choice experiment featured a special subset of the multi-attribute that varies only in price, a format consistent with contingent valuation studies. That is, each potential user is presented with the same attributes (or level of services) with price varying across different users. With these options in mind, the CAPM research team conducted three focus groups and held a workshop to finalize the design. The research team informed the focus group participants that their feedback would be used to develop a survey of library users and provided them with incentives for their participation.

The first focus group consisted of JHU undergraduate students. The research team demonstrated the possible features of a CAPM system and, subsequently, provided participants with written, sample choice experiment formats. These formats offered several multi-attribute choice panels in which alternatives differed by the presence or absence of digital (scanned) images and full text search, and varying levels of delivery time and price as attributes. Additionally, the research team presented a second format with all CAPM system attributes at a specified price. In this second format, the service attributes are the same, but prices are varied across different users. The first focus group participants unanimously expressed that they preferred the first format offering varying services with varying costs. The research team then focused on the issue of assigning dollar values for library services. The first focus group indicated that they were comfortable with the notion of trading money in exchange for CAPM services. One participant noted that JHU students are well aware that educational services come at a cost.

The second focus group comprised JHU faculty. The research team presented a set of sample choices that reflected the comments from the first focus group. The faculty group commented on the visual layout of the proposed survey and conveyed their experiences with requesting items from Moravia Park. The faculty group also preferred the multi-attribute choices (i.e., comparison of different levels of services with varying prices) and expressed comfort with the notion of paying for library services.

The third and final focus group involved JHU graduate students and staff. The third group confirmed the findings from the first two focus groups. This feedback was augmented by the findings from a workshop held at JHU.
A group of economists, librarians and information scientists gathered for a workshop with the objective of designing an appropriate multi-attribute, stated-preference choice experiment. This workshop occurred between the first and second focus groups. The workshop participants provided feedback and advice regarding the survey choices, layout and process. Professor V. Kerry Smith from North Carolina State University, an economist specializing in non-market valuation, provided a particularly helpful observation. He asserted that the survey should focus on CAPM services, rather than its technology. The first focus group participants (undergraduate students) were particularly interested in the technology underlying the CAPM system. Professor Smith expressed concerns that untried technology may undermine the credibility (or at least the perception of credibility) for the choices for the experiment (e.g., users might not believe possible delivery times). Furthermore, users might base their choices on appreciation or fascination for cutting-edge technology rather than the services offered by CAPM. Consequently, the choice experiment survey was designed to focus on the possible services, rather than the technology, associated with CAPM.

With the feedback from the focus groups and workshop participants, the CAPM research team confirmed that a multi-attribute choice format, offering a comparison of different levels of services for varying prices, was most appropriate. The CAPM team then proceeded to develop experiments, in the form of surveys, based on these choices. Surveys that reflect multi-attribute choices often begin with introductory questions that help focus the users’ attention. For the CAPM survey, the first question related to the users' affiliation (e.g., faculty, staff, student). The subsequent questions focused on general library usage and familiarity with Moravia Park. Users were then provided with an overview of Moravia Park and/or the CAPM system (without focusing on the technology). These introductory questions helped focus the users' attention on the possible services that would be offered by CAPM.

The next section of the survey presented the choice experiments. Varying the levels of attributes or services in a rational manner constitutes a proper experiment. There is a fairly extensive literature regarding the statistically efficient design of the attribute or service levels that will allow identification of marginal effects while requiring as few choices as possible from the users. The crux of good experimental design rests upon the appropriate choice of variation among the levels of attributes or services. As mentioned previously, there were four attributes for the CAPM system. The research team chose the following levels for these attributes:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Range of Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital (Scanned) Image</td>
<td>0, 1</td>
</tr>
<tr>
<td>Full Text Search</td>
<td>0, 1</td>
</tr>
<tr>
<td>Average Delivery Time</td>
<td>6 hours, 2 hours, 20 minutes, 30 seconds</td>
</tr>
<tr>
<td>Price per Semester</td>
<td>$0, $15, $45, $70, $110</td>
</tr>
</tbody>
</table>

The first two attributes were considered binary since they are either present or absent. The research team decided the combination of no digital (scanned) image and full-text search was unrealistic. The full-text search assumes the existence of text that has been processed.
from OCR of a digital image. The research team assigned a six-hour delivery time and $0 cost to the present system. The other levels for both average delivery time and prices per semester reflect the various possibilities associated with a fully operational CAPM system. It is important to note that "delivery" for the CAPM system describes the transmission of digital images over the network. Depending on whether an item has been scanned previously, the delivery time can vary.

Within the survey, each user faced three alternate systems that reflected different levels of service. A sample choice panel is provided in the following table:

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Current System</th>
<th>System A</th>
<th>System B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Wait Time</td>
<td>6 hours</td>
<td>2 hours</td>
<td>20 minutes</td>
</tr>
<tr>
<td>Digital Image</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Full Text Search</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>User Cost (per semester)</td>
<td>$0</td>
<td>$70</td>
<td>$110</td>
</tr>
</tbody>
</table>

In this specific case, the sample choice examines the user's valuation of average delivery time. Each user faced ten of these choices, with each choice panel offering different levels of services. The number of alternatives within each panel and the number of choices reflected current practices in the literature. Some combinations of attributes and prices were not considered. For example, aside from the current system, no system was assigned a six-hour delivery time or a zero price. Given the assumption that the combination of no digital image and full-text search was not possible, there were three possible digital image and full text search combinations (i.e., no digital image, no full text search; digital image, no full text search; and digital image, full text search). Combined with four prices, and three delivery times, there were $3 \times 4 \times 3 = 36$ alternative possible systems. Converting these possibilities into choice panels such as the one provided in Table 2 yielded 630 possible arrangements of the current system juxtaposed with two alternative systems from the aforementioned 36 alternative systems. To deal with this large number of possible panels, the assignment of alternative systems into panels and the grouping of 10 choice panels (questions) were generated randomly.

With this design, a graduate student at the University of Colorado at Boulder developed a database-driven, web-based survey. This approach reduced the costs associated with the survey by facilitating data entry and analysis, and eliminating the need for postage. Even though the web-based survey made it easier to reach audiences beyond JHU, and the CAPM system has the potential to be used by individuals beyond JHU, the research team elected to focus on JHU library users as the relevant population for the survey. This choice also reflected the desire for tractability and manageability of the effort. The research team obtained a database of email addresses for JHU affiliates. There were some important omissions from this database, including a portion of the part-time students at JHU.
Nonetheless, the email database provided a reasonably representative cross-section of potential CAPM users at JHU.

Before implementing the survey with a larger group, the research team tested the survey through a pretest of three hundred individuals. This pretest sample was partitioned into two groups. One group was asked to complete the survey without incentive. The other group was offered an incentive: anyone who completed the survey would be entered into a lottery drawing for a gift certificate at a local, upscale restaurant. This pretest provided evidence that an incentive would foster greater participation, and feedback that led to final revisions of the web-based survey.

For the final survey, the research team elected to use both email and (campus) paper mail contacts. The goal for the survey was to have at least five hundred responses. With this goal in mind, the research team contacted two thousand randomly selected JHU affiliates, in direct proportion to the percentage of faculty, graduate students, undergraduate students and staff at JHU. The research team implemented a few measures in an effort to increase the response rate. Each of the participants was contacted with three email and two campus mail letters over a period of days. These contact letters were personalized to include the individual's name (from the JHU database) and addressed from the CAPM research team leader. Finally, each person who completed the survey was entered into a lottery for a $500 gift certificate from a local travel agency.

Six hundred and three individuals completed the web-based choice survey, resulting in an unadjusted response rate of 30%. Table 3 summarizes the sampling and unadjusted response rate information by user group.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Sampling &amp; Response Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Population Proportion</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Undergraduate</td>
<td>0.39</td>
</tr>
<tr>
<td>Graduate</td>
<td>0.14</td>
</tr>
<tr>
<td>Faculty</td>
<td>0.27</td>
</tr>
<tr>
<td>Staff</td>
<td>0.20</td>
</tr>
<tr>
<td>Overall</td>
<td>1.0</td>
</tr>
</tbody>
</table>

The results from this survey indicate a hypothetical willingness to pay approximately $63 per semester for services associated with CAPM. This figure was consistent regardless of affiliation (e.g., undergraduate student vs. faculty) or familiarity (or unfamiliarity) with Moravia Park. This valuation of benefits compares favorably with the potential costs associated with a CAPM implementation (Lorie 2001). While the project team calculated willingness to pay, it was not necessary. Instead of comparing exchanges of money with varying levels of services, it would have been possible to compare exchanges of time or different services (e.g., would users prefer to spend funds on electronic acquisitions rather than enhanced access through CAPM). Flores (2001) provides greater detail regarding the benefits analysis and results.
Conclusions

The economic analysis of the CAPM system has been instrumental in encouraging JHU to continue with the project. Perhaps more importantly, it has encouraged other institutions to consider adoption or implementation of CAPM. The CAPM research team leader provided two presentations at American Library Association conferences, the first in 2000 and the second in 2002. In 2000, conference participants were interested, but somewhat skeptical, of the system. In 2002, several conference participants expressed specific interest regarding implementation, citing the economic analysis as a key factor in their interest.

While the specific results of this analysis are important, it is even more important to consider the applicability and utility of adopting the evaluation framework for other digital library services. With increasing expectations from users, greater accountability, and rising costs, the need for evaluation has never been more pressing. The CAPM Project experience provides evidence that a carefully designed and implemented multi-attribute, stated-preference analysis can offer invaluable feedback from users.

References


Notes


(HTML coding for navigational link to editorial corrected 8/31/05.)

Copyright © Sayeed Choudhury, Benjamin Hobbs, Mark Lorie, and Nicholas Flores