

Quantitative Study of the Value of Research Libraries: A Foundation for the Evaluation of Digital Libraries

Paul B. Kantor and Tefko Saracevic

LIS and Rutgers Distributed Laboratory for Digital Libraries
School of Communication, Information and Library Studies Rutgers University, New Brunswick, NJ
{kantor, tefko}@diglib.rutgers.edu

Abstract

In anticipation of the explosive growth of digital libraries, a complex study was undertaken seeking to evaluate 21 diverse services at 5 major academic research libraries. This work stands as a model for evaluation of digital libraries, through its focus on both the costs of operations and the impacts of the services that those operations provide. The data have been analyzed using both statistical methods and methods of Data Envelopment Analysis. The results of the study, which are presented in detail, serve to demonstrate that a cross-functional approach to library services is feasible. They also highlight a new measure of impact, which is a weighted logarithmic combination of the amount of time that users spend interacting with the service, combined with a Likert-scale indication of the value of that service, in relation to the time expended. The measure derived, incorporating simple information obtainable from the user, together with information which is readily available in server/client logs, provides an excellent foundation for transferring these measurement principles to the Digital Library environment

INTRODUCTION

This report is one result of a 15-month long project to study the costs and value of library functions at five major research libraries. A total of 21 services or service aspects were studied. Numerous measures of the importance or benefit of the service to the users were studied together to lay a foundation for the development of an economically valid scale for assessing the impact of library services. Further, more than 500 interviews were transcribed, and through detailed content analysis an Empirical Taxonomy (reported elsewhere: Saracevic & Kantor, 1997) was developed for classifying the contexts and the value to users of library services. Cost estimates, using functional cost analysis, were developed for all the services. The principles of Data Envelopment Analysis have been adapted to this situation, and illustrated using representative measures of library impact.

This study has definitely established not only the characteristics of the numerous services studied, but also the fact that a single uniform instrument can be applied to study many diverse services at many libraries. The long-term goal of this research is the development of a general metrology for library benefits. Substantial progress has been made, and yet more remains to be done. The study has identified specific conceptual problems that arise when the goal is to extract economically useful information from interview data, and points the way to further methods that will resolve these problems.

Overall, we have found that users of services at research libraries (1) value these services very highly, (2) have very little experience purchasing information services, (3) value the library more highly than other university-supplied services, (4) do not assign dollar estimates to the value of services which are commensurate with the cost of the services, and (5) can articulate, with proper questioning, the context and purposes which bring them to use library services.

PROBLEM AND OBJECTIVES

Research libraries and their supporting institutions are increasingly concerned about costs incurred and values received. This concern grows from difficult decisions on allocation of limited resources and the bewildering array of modern technology-based resources and services available to libraries. Libraries provide an ever

increasing array of services. But the library is an internal service supported by the institution as a whole, in aid of its overall goals and objectives. Ultimately, the critical question that must be answered by institutional managers is:

Does the sum total of value flowing from the library justify our cost in maintaining it?

This most difficult problem leads, in turn, to many other ones. Studying the cost of university libraries raises a host of interesting methodological challenges. Adding the study of value, as perceived and/or defined by users, increases the methodological challenge many fold. These concern three aspects: the definition and determination (that is, the model and method) of the cost of library services; definition and determination of value of services to users; and finally combining in some way both cost and value together.

The goal of the study reported here was to develop models and methods for studying the cost and value of library services in a way that can be pragmatically generalized and applied as libraries radically change the technologies by which they provide services. We sought methods for gathering information on the cost and value of library services which concentrate on factors that are defined across technical modalities such as: user assessment of benefits, time spent using the service, and dollar cost of the service.

The objectives of the overall study were to:

1. Define and apply methods for obtaining costs for several library services.
2. Derive an empirical taxonomy of values for these services based on users' assessments.
3. Provide methods for combining cost and value data.
4. Provide a detailed description and manual that will allow for replication of these types of studies.

To achieve these goals and objectives required an empirical study involving collection of a large amount of data from several libraries and a number of different services. Five libraries, 21 services, and over 500 user interviews were involved, making this one of the largest, if not the largest study of cost and value of library services. The study incorporated development of appropriate models and methods for study of cost and value, extensive validation of these with data from actual services and situations, and extensive documentation of these efforts. Results of the work driven by the 2nd and 4th objectives have been reported elsewhere (Saracevic & Kantor, 1997). This paper is the first conference report on the issues addressed in objectives 1 and 3. It draws on methods and results reported in (Kantor et. al. 1994; d'Esposito-Wachtmann et al. 1994a,b)

The non-profit setting

The critical question posed at the outset (Does the sum total of value flowing from the library justify its cost?) is difficult enough to resolve in the corporate world, where there is at least one agreed upon overall measure of corporate progress: the net profits returned to shareholders. While this choice of a measure is subject to criticism from many directions, and is sometimes blamed for the failures of American industry to carry out long-term strategies aimed at developing market share, it does at least provide some starting point for the study of the role of libraries. The major research university, on the other hand is a quintessential example of a non-profit organization. Unlike hospitals, universities do not deal with a series of well-defined incidents or cases, which might be studied one by one to assess the impact of the library. Universities produce essentially intangible products such as well educated students and cutting edge research. Thus, while financial soundness and fiscal responsibility are essential in the operation of a university, balancing the budget does not reflect progress towards those major goals. Even when a university broadens its goal statement to include a desire to have an impact on surrounding communities, and on the national cultural or economic profile, these enlarged goals also do not point the way to clear internal measures of progress.

Scholarly and Practical Importance

The question of value is attractive from a scholarly perspective. The issue from the scholarly perspective is How can we reasonably define and measure the contribution of the library (whose deliverable products are themselves intangible) to the overall goals and objectives of the university as a whole, whose goals and deliverables are even less tangible?".

Aside from this scholarly interest there is a pressing practical reason to be concerned about measuring the value of the library. The library has stood as a unique service organization in the university setting. Enshrined in a major campus building, it serves as a natural community focus for scholars, a point of pride for the entire community, and its very bulk and contents have seemed in themselves to answer the question Why do you cost so much?. All of this is being changed with the rapid development of computers and telecommunication systems which link the scholars and the students at their desks and laboratory benches to the metadocuments (catalogs, bibliographies, etc.) of the library on campus, of libraries around the world, and increasingly to the documents themselves maintained locally or elsewhere. With this in mind we have tried to include in our study examples of the "metadocumentary" services, such as enhancements to the online public access catalog, as well as services which provide access to information and to documents themselves. All of this paves the way for the transition to measuring digital libraries in the same terms that we measure conventional libraries.

As the richness of a library is increasingly measured by the range of materials which students and scholars can reach, rather than by the physical possession of those materials, it is essential that we develop measures that assess, as well as we can, the value of present library services to the institution. With such measures in hand, libraries will be prepared to move through the transition, documenting, as they go, that a decrease in the number of serials titles held or the number of bound volumes on the shelves, need not represent a decrease in the library's impact on the university. In fact, the best measures of impact would show the increase in value as access to materials is made more transparent for students and scholars.

Measurement Focussed on Patrons

As a university does not conceive of itself in terms of manufacturing divisions, or product lines, there is no point in the institutional framework to which to address the question "How much does the library benefit the institution?". Rather, we believe, the correct point of inquiry is the individuals who make use of the services of the library. They generally do not conceive of their use of the library in terms of impact on the institution at large. Rather, they see themselves, at the moment of use, as engaged in some particular task or project whose goals are more or less in line with overall goals of the institution.

Setting aside tasks and projects that are entirely personal in nature (hobbies, personal health concerns, preparation of tax returns, etc.) we can anticipate that students will primarily see their tasks in terms of completion of course work or research, in progressing towards an approved degree to be awarded by the university. Scholars on the other hand engage in a variety of tasks broadly defined as service or administrative tasks (such as verifying the credentials of a colleague who is being considered for promotion), teaching (which involves maintaining current awareness of pedagogical developments, as well as the preparation of specific lectures, assignments, exercises, laboratory projects, and so forth) and scholarship or research.

In this situation we decided that the most effective approach to assessing the value of the library is to focus attention on the specific task or project that brings the user to the library, and to ask questions about value, benefit and importance in the specific context of that use. As we discuss at length in (Kantor & Saracevic, 1994), it is not an easy matter to ensure that respondents maintain this focus even during a brief ten-minute interview.

Ideal Economic Perspective

From the economic point of view one would build up the impact of the library on the institution by combining all of the instances in which the library contributes value to some task or project, weighting each such task or project in proportion to its own contribution to the goals and objectives of the university at large. This is, at present, only a broad conceptual structure, which can not be implemented. Universities have extreme difficulty in assigning any kind of relative weight or importance to their various missions, generally adopting (at least for the public) the stance that all components of the university mission are essential and therefore equally important. Similarly, it would be quite difficult to assess the importance of any single task or project as a contribution to any of the specific objectives of the university. However, even if this embedding of the results of a study such as ours into a larger economic picture of the institution can not be completed, we believe that there are techniques which will permit meaningful comparison of distinct modes of library operation in this setting.

The idea behind our analysis is to take as a kind of constant or invariant the ability of the users of the library to assess the value of the library to their own tasks or projects. In other words, if the library, over a span of some ten years, is increasing its value to the tasks and projects of its users, then the natural presumption is that it is increasing its value to the institution. That assumption could only be attached by a concrete demonstration that in some way the tasks and project of the users have evolved to be less in line with the goals of the university during the same period. Therefore, we believe that it is of importance to gain a deeper understanding of the value of the library to tasks and projects of the users.

The costs of service

We study the costs of library services from the point of view of unit cost, or functional cost analysis. Unit cost analysis in production industries seeks to assess the total cost of producing a single object, such as a toothpick or a computer. The analogous concept for service industries involves defining specific units of service. Thus, the task is to assess the costs per service unit delivered. The specific point of view taken here does not address economic subtleties such as economies of scope or economies of scale. Rather it follows established accounting procedures that allocate all internal costs not directly associated with any particular service, to the range of services provided.

Scales for Assessing Value

We have studied the value to users in terms of scale questions on an interview instrument. In these scale questions users are asked to represent their response with regard to something like the importance of a service on a 7-point scale. Such scales are generally recognized as having ordinal validity, but it is questionable whether they can be treated directly as interval scales of the type needed for most economic modeling. The specific set of questions that we used in the later stages of our study were refined through earlier preliminary studies, and still represent, in our judgment, a work in progress. Details of the questions themselves and of the potential that they offer for the development of one or more abstract scales of library value are given in (Kantor & Saracevic, 1994; d'Esposito-Wachtmann et al. 1994a,b).

Overview of the Libraries and Services Studied

The study involved five large academic libraries in research oriented universities, which can be considered both academic and research libraries. The universities participating in the study are shown in Table 1

<p>Table 1 Participating Universities.</p> <ul style="list-style-type: none">• <u>University of Maryland</u>: A publicly supported research university with approximately 19,700 students.• <u>New York University</u>: A privately supported research university with approximately 21,400 students.• <u>Princeton University</u>: A privately supported research university with approximately 6,400 students.• <u>Columbia University</u>: A privately supported research university with approximately 14,700 students.• <u>Rutgers University</u>: A publicly supported research university with approximately 33,100 students. <p>* Number of students refers to full time students on the campus where the library is located.</p>

In this paper, the numbering of libraries from 1 to 5 does NOT correspond to the order of listed universities. We do this to preserve the anonymity of services studied.

The study involved two time frames: Fall of 1993 and Spring 1994, with preliminary screening interviews or contact cards, followed by telephone interviews whose details are given in (Kantor & Saracevic, 1994). A general template questionnaire was developed, and adjusted to fit specific characteristic of each service studied. Thus, the questionnaires for different services were specific to the services, but followed a general model that supported combined analysis. Interviews were conducted by trained interviewers following the questionnaires and instructions. The manual, instruments, and instructions for training of interviewers are detailed in (d'Esposito-Wachtmann et al.,1994a,b).

Table 2 shows the relations among the services and the libraries and Table 3 elaborates the services studied, with very brief descriptions. More detailed descriptions of each service are given in (Kantor & Saracevic, 1994) Section 3.2. The Codes given in the first column of this table are referenced in the results of the Cost Analysis.

Table 2. <u>SERVICES STUDIED</u>	
<u>LIBRARY 1</u>	
● Reference Services	Information desk In-person reference Reference consultation
● Materials Delivery Service	
<u>LIBRARY 2</u>	
● Art and architecture library - the collection	
● Biology library - electronic reference services	
● Psychology library	
<u>LIBRARY 3</u>	
● Enhanced online catalog	- on-site - phone
● Undergraduate reserve service	- students - faculty
● Science document delivery service	
● Carl Uncover pilot program	
<u>LIBRARY 4</u>	
● Electronic reference services	
● Music and media center	- students - faculty
● Interlibrary loan service	
<u>LIBRARY 5</u>	
● Patents service	
● Reference	
● Automated reference	

SUBJECTS in SAMPLE

- Faculty
- Graduate Students
- Undergraduate Students
- Non-institution affiliated users

Table 3. Table of Services Studied.		
Code	Lib.Service	Brief Description
R5	L1.S1.1	Information Desk
R8	L1.S1.2	Reference Desk, Branch 1
R4	L1.S1.3	Reference Desk, Branch 2
R7	L1.S1.4	Reference Consultation Service
D1	L1.S2	Materials Delivery Service (MDS)
L1	L2.S1	Art and Archeology Library
R6	L2.S2	Biology Library - Electronic Reference
R9	L2.S3	Psychology Library - Reference

C1	L3.S1	Enhanced Online Catalog Service
C2	L3.S1.V2	Remote Users
V3	L3.S2	Undergraduate Reserve Service
V4	L3.S2.V2	Undergraduate Reserve Service: Faculty
NotCosted	L3.S3.1	Document Delivery Service
D2	L3.S3.2	Carl UNCOVER
R2	L4.S1	Electronic Reference Resources
X1	L4.S2	Interlibrary Loan Service
V1	L4.S3	Music and Media Center
V2	L4.S3.V2	Music and Media Center: Faculty
R3	L5.S1	Automated Reference Service
R1	L5.S2	Traditional Reference Service
L2	L5.S3	Patents Service

COST ANALYSIS

Our determination of library costs follows general principles for unit cost accounting. The application of these to the library setting has been set out in considerable detail elsewhere (Kantor 1986, 1989) and will not be reviewed here. It is not difficult to identify the direct costs associated with a particular service, such as the supplies consumed or the labor of people who work on no other service. It is somewhat more difficult when individuals work on several services, but well established techniques can be used to allocate their time among those services. Essentially, there are only two difficult problems in the cost allocation: allocation of salaries, and allocation of the cost of shared resources. The techniques used in this study were adapted, with permission, from the Tantalus Inc. FUNCOST software for library cost analysis and are summarized here.

Methods for Cost Analysis

To begin with, costs can be divided into start-up costs and ongoing costs. In this project all services had been established some time before the analysis was done, and only ongoing costs are considered. Costs are further divided into fixed costs and variable costs. The fixed costs are those costs necessary to provide the service at all. For example, a circulating collection must contain some books. The variable costs are those costs which are approximately proportional to the volume of services. In the accounting model of cost analysis used here, the average costs are reported. This is similar to treating all costs as variable costs. Thus the unit costs reported here, while they provide a full account of expenditures, cannot be used as reliable predictors of cost change for all other situations. For example, if the volume of service increases moderately, the corresponding increase of cost, at the same level of service quality, will probably be less than the figure reported here. If, however, the present system is operating close to capacity, an increase in volume is likely to produce a noticeable decrease in service quality, unless there is some compensating investment in equipment, personnel, or both. Such a quantum jump in capability would result in an increase in the average cost of service.

Economies of scale, which may be either positive or negative, represent deviations from the straight-line dependence of variable costs on volume of service. For example, with positive economies of scale, the average unit cost decreases as the volume of service increases at constant service quality. Similarly, with positive economies of scope, the total cost of providing several different types of service will be less than the sum of the costs of providing them separately. It seems reasonable to suppose that economies of both scale and scope are present in library operations, and in the electronic versions of information service that are an increasing factor in library operations. However, this study is, by design, limited to consideration of services in isolation, and at a few libraries. Thus it is not possible to model either economies of scale or economies of scope.

To sum up, our model calculates, for each service, the average unit cost of service at the present level of operation. All shared resources are pro-rated, using appropriate methods, to the one service under consideration, and to all other services. This method probably underestimates the cost of providing any particular service on a stand-alone basis. This shortcoming is not of interest to the community of research libraries, who are the primary audience for this study, as they do not offer services in isolation.

Costs of equipment, in industry, are normally dealt with by reference to depreciation tables, or amortization schemes, which are, in turn, developed by industry in response to the tax laws. Since all of the institutions studied are non-profit, this information is not available. We have assumed throughout that computer equipment is to be expensed over a three or a five year period, depending on the likely reasonable life. We note that the reasonable life of computer equipment today is determined more by considerations of obsolescence than by considerations of maintenance and useful working life.

Labor costs are allocated according to each worker's estimate of the proportion of time allocated to that service. Time related to professional development is simply dropped, effectively increasing the cost of all other activities as if that share had been distributed over it. In other words, the cost per unit of this person's work on end-products is increased. In this particular study we did not analyze all of the services and functions provided by each participating library. This means that we do not know the correct base on which to allocate the administrative time, particularly committee work, among the several services. After some efforts to reconstruct this data by extended conversations with the libraries involved we concluded that it would not be possible to develop uniformly reliable and comparable measures across the several libraries. Examples of the problems are (1) library administrators who are involved in supervising multiple services and who would have to allocate their time against all of those services, requiring a much more detailed cost study than has been undertaken here and (2) library professionals who spend a significant amount of time in committee meetings but have no natural basis for allocating that committee time among the several tasks on which they spend their time. Thus, for practical purposes, the calculation of labor cost concludes with the distribution of the annual salary to specific services, in the proportions indicated by the staff, and we do not "burden" the reported costs with administrative costs.

Computer costs are usually distributed according to some reasonable measure of the use of the computer. Suitable candidates for this measure include: the number of ports dedicated to a service; the number of CPU cycles (or CPU seconds) consumed by that service, or the number of terminals provided for the public to access that service. In all of the services studied here, the number of terminals was used as the basis for cost assignment.

We believe, based on our examination of the raw data, that there are substantial variations to be observed in the ratio of administrative costs for a service to the other direct costs of the service. We are not in a position to separate that variation into its two most natural components: variation due to the nature of the service being managed, and variation due to the nature of management procedures at the library being studied. These issues are of importance in connection with the total quality movement in library studies, but are beyond the scope of this study.

In addition to the problem of administrative costs there is the problem of allocating the cost of the collection, and of the organization of that collection, in determining the functional or unit costs. In a full scale study of a library this is done by forming all of those costs into an overhead pool and allocating them among several services in proportion to the degree to which those services make use of the books and other materials in the collection. In this study we do not have the information necessary to complete this. For most of the services it does not seem to be an issue. For reference services in particular, we do not include the costs of the specific collection of reference materials maintained, and we set aside any consideration of the use that reference service makes of the collection as a whole.

However, for one of our specific services it is not possible to sidestep this question. That is the use of the non-circulating collection at an Art & Architecture Library. In this case we have accepted the annual book purchase cost, and the assigned share of central processing costs, as costs of the service provided. In the same spirit, we have included the cost of the reserve materials at an undergraduate reading room. However, it was not feasible to use this approach in determining the costs of the audio-visual center at a

different library. Similarly, the cost of developing the catalog records, which are accessed by the online system, have not been included.

Fringe costs were systematically excluded, to make the comparison more uniform. We did, however, use the actual labor costs. There are variations in wage scale, as two of the libraries are in New York City, two are in New Jersey within 50 miles of New York, and one is in suburban Washington DC.

Cost Data Results

Bearing these caveats in mind, we now turn to a summary of the costs of the specific services (Table 4). To assure the confidentiality of the libraries, we have separated the numbering of the libraries from the naming of the libraries, as described earlier.

Table 4 Reported Costs of Services, labelled by general category of service. Services are presented in increasing order of unit cost. (*D=Materials Delivery. C=Catalog or online service. R=Reference Service. V=Reserve Collection. L=Entire Collection. X=Other service. See more detailed descriptions of the services in Volume 2: Process.*)

Cost Code	Lib.	Service	Brief Description
N/A	XX	L3.S3.1	Document Delivery Service
\$0.70	C1	L3.S1	Enhanced Online Catalog Service
0.70	C2	L3.S1.V2	Remote Users
1.16	R1	L5.S2	Traditional Reference Service
1.88	D1	L1.S2	Materials Delivery Service (MDS)
2.00	R2	L4.S1	Electronic Reference Resources
2.65	V2	L4.S3.V2	Music and Media Center: Faculty
2.65	V1	L4.S3	Music and Media Center
6.12	L1	L2.S1	Art and Archeology Library
6.87	R3	L5.S1	Automated Reference Service
8.18	R4	L1.S1.3	Reference Desk, Branch 2
8.59	V3	L3.S2	Undergraduate Reserve Service
8.59	V4	L3.S2.V2	Undergraduate Reserve Service: Faculty
12.30	X1	L4.S2	Interlibrary Loan Service
14.62	L2	L5.S3	Patents
18.35	R5	L1.S1.1	Information Desk
18.80	R7	L1.S1.4	Reference Consultation Service
20.35	R6	L2.S2	Biology Library - Electronic Reference
33.36	R8	L1.S1.2	Reference Desk, Branch 1
35.52	R9	L2.S3	Psychology Library - Reference
36.13	D2	L3.S3.2	Carl UNCOVER

We see that costs range from a low of less than a dollar, typically for the use of some computer-supported service that is heavily trafficked (and for which we have not included the costs of developing the underlying database), to a high of over thirty dollars, for some classes of reference and materials delivery services. These figures must be regarded as approximate, since there are variations in the degree to which certain administrative costs have been included, there are arbitrary decisions made in distributing the cost of shared equipment, and they do not include fringe costs. However, we are confident that they accurately represent the range of costs for the services considered.

We note that the range of costs for reference services is enormous: from a low of \$1.60 to a high of \$35.52. The low end is surprising, while the high end is consistent with the results of other detailed studies of reference costs in the context of full library studies. This figure is, presumably, completely unsuspected by patrons who assign (in interviews) a value of \$5-\$10 to a library service event. Of course the average cost

of this service is hard to control, since the library sets out the service, but the patrons do or do not make use of it.

Similarly, the two materials delivery services vary enormously in cost. Interestingly, the two cases in which we studied an entire collection fall into the middle of the range, with the specialized patent operation being somewhat more expensive than the (equally specialized) art and architecture setting. As has been determined in other studies, interlibrary loan service costs far less than the cost of buying and processing a book, although we note that this includes only the costs experienced by the borrowing library.

More detailed confidential reports have been prepared for the participating libraries, and those will permit further discussion of the accuracy and comparability of the numbers presented here. Ultimately the participating libraries, which have been assured confidentiality with regard to cost and performance data may or may not decide to meet and share that information.

COMBINING COST AND VALUE DATA

Impact Data

We have explored the potential of combining two measures in a non-standard way. In particular, we have information on the assigned dollar value of the service, on the length of time spent using the service, and on the question "was the value worth the time?". We first explored the estimates of dollar value by itself. The distribution of this variable shows peaks at convenient numbers, and is highly skewed, with a concentration at very low values, and a high mean, influenced by a few outliers. We trimmed the variable by assuming that estimates in the range of \$4,000 to \$5,000 could be economically realistic, but that estimates of \$50,000 could not be. We then explored the variation of the mean of the dollar value with the type of service. No statistically significant variation was found.

We found that none of numerous scale values (see K&S 1994) intended to assess the impact of libraries yielded significant dependence on the service type. Similarly, analysis of the time spent yielded no significant dependence. We next transformed the time spent to a logarithmic scale (based on natural logarithms), calling the new variable LogTime. This also did not show any significant variation with the service. However, the **time** spent by users at a library does represent some kind of an economic decision, and we sought to weight this new variable, the time spent by users using the information, based on information that they provided about the relation between the benefit to them and the time spent using the service.

We defined a composite variable:

$$V = \log(\text{time}) * (B-1)/6.$$

Here B is the scale value reported for "is the benefit worth the time". The highest value of B is 7, in which case the new weighted variable has the same value as the logarithm of the time spent. But for the lowest value of B (B=1), this weighted variable vanishes. The time spent using the service is measured in minutes, and is always greater than 1 minute, so this new variable is thus either positive or zero.

Analysis of variance revealed that this new variable is just barely able to resolve differences among the services. Specifically, the data let us reject the hypothesis that it has the same mean value for all the services. But further analysis (Scheffe "post-hoc" analysis) reveals that only the highest of the values found for the mean of the weighted logarithm of the time is significantly different from some of the other values. And, at that, it is significantly different only from the four lowest values.

Thus the measure **V** which is used in the remainder of this discussion is still only a weak discriminator of the value of the services. Nonetheless, it represents the best that we have been able to extract from the available data, and will serve to illustrate the concepts of Data Envelopment Analysis.

Combining Cost and Impact Data

To illustrate how one can address both cost and value simultaneously we assemble here the data on the weighted log-time value measure (**V**), and the cost data given above (Table 5). In order to produce a usable graphical representation we represent the cost on a logarithmic scale, by calculating LogCost.

Table 5. Weighted Logarithm of the Time Spent Using the Service combined with cost per unit of service. The services are defined in more detail in Table 1.2. (*D=Materials Delivery. C=Catalog or online service. R=Reference Service. V=Reserve Collection. L=Entire Collection. X=Other service.*)

Code	Mean	StdDev	N	Cost	LogCost
C1	1.4089	1.1017	24	0.70	-0.34
C2	1.8783	1.0654	30	0.70	-0.34
R1	1.4113	0.7576	47	1.16	0.15
D1	1.1145	0.8669	52	1.88	0.63
R2	1.8642	0.9791	47	2.00	0.69
V1	1.3795	1.3662	26	2.65	0.98
V2	2.1946	1.5481	7	2.65	0.98
L1	1.5764	1.1947	49	6.12	1.81
R3	1.289	1.3758	48	6.87	1.93
R4	1.0997	0.6624	23	8.18	2.10
V3	1.3712	1.0751	7	8.59	2.15
V4	1.1797	0.8765	28	8.59	2.15
X1	1.3037	0.9727	14	2.30	2.51
L2	1.454	1.3965	17	4.62	2.68
R5	0.8463	0.5615	10	18.35	2.91
R6	1.4841	1.094	41	20.35	2.93
R7	0.9672	1.3678	2	18.80	2.93
R8	1.784	1.6408	7	33.36	3.51
R9	1.8322	1.1015	25	35.52	3.57
D2	1.4433	1.0171	10	36.13	3.59

One of the services could not be independently costed, and was assigned the cost of "0", which cannot be transformed onto a logarithmic scale, so it is dropped from this analysis. Note also that when the unit of cost is taken to be the dollar, values less than \$1.00 are represented by negative values of the logarithm of the cost. This does not represent a negative cost, but is just a reflection of the arbitrary unit of currency.

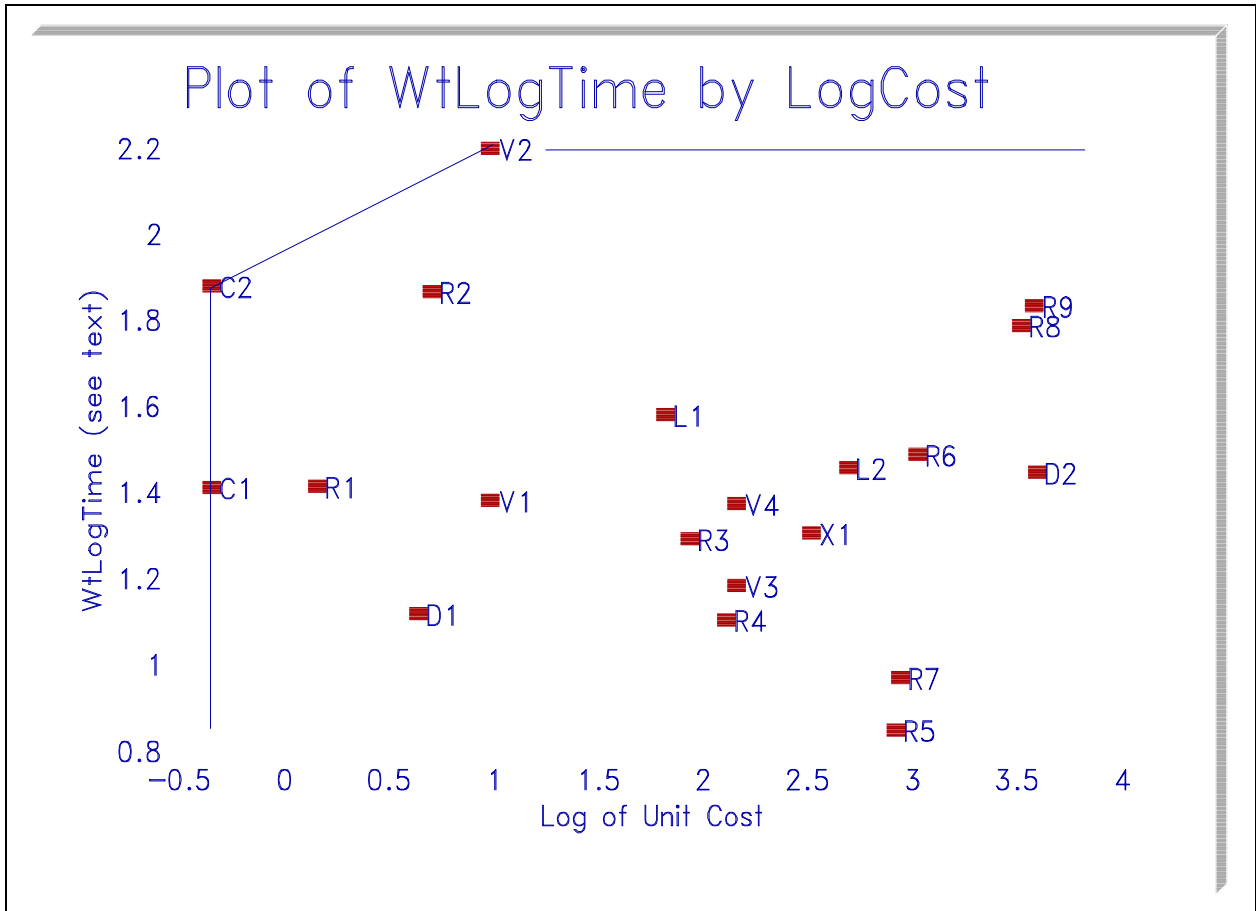


Figure 1. A Joint Plot of the Presumed Measure of Impact and the Determined Cost of a Unit of Service. Points closer to the upper left hand corner represent greater impact for the money spent on the service.

The two variables can be plotted together, in a graph (Figure 1) on which each point represents one specific service, the horizontal axis represents the (logarithm of) the unit cost of the service, and the vertical axis represents the new composite measure of value **V**.

Data Envelopment Analysis

In data envelopment analysis (Ahn et al., 1988, Banker et al., 1984, Charnes et al., 1978, Seiford, 1990) service units, or functions, are represented by points on a graph such as the one shown in Figure 1. One axis represents the cost of the service, and the other represents the impact. Such an array of data would conventionally be analyzed by some type of regression analysis. That produces the “average line”. For data such as shown here the average line does not tell us very much, as the individual services are widely scattered.

Under data envelopment analysis we concentrate on the boundary defined by these points. In this case it is the line determined by the two leftmost points, labeled C1 and C2, and the highest point, labelled V2. These represent, among them, the three services which cannot be beaten. All other services have either higher unit costs, or lower impact (as measured by our surrogate) or both. In a sense this is a two-dimensional variant of the familiar notion of "benchmarking" which seeks to identify best practice for any particular service or function. Since different services are being considered together, the benchmark cannot be reduced to a single number, but must consider the cost and the impact separately. If we were confident in the validity of our measure we could in fact proceed to calculate the degree to which each other service

is less than optimal. In the present, exploratory situation, such a display of analytical prowess would be unjustified.

We note a few interesting features of the DEA plot. First, the catalog services are clearly least expensive. This is expected, particularly as we have not included the costs of developing the database to which it provides access. Two cataloging services appear twice with the higher impact value (C2) being reported by users who were interviewed by phone. Thus it has a greater positive impact for those who use it from remote sites than for those who use it at the library (C1). This suggests that either it is used longer, or that the user adjusts the estimate of benefit to compensate for the effort required to go to the library. The reserve room services "V" have costs in the middle range, and impacts ranging from the highest to among the lowest. In fact, V1 and V2 correspond respectively to student and faculty estimates for the same service. So do V3 and V4. Thus, in each case, the impact measure assigned by faculty exceeds that assigned by students. This suggests that the benefit to the faculty of the reserve services is greater than the benefit to the students on a single visit. We must also note that faculty members typically have a single long interaction with the reserve room, and we have not separately costed the fulfillment of that interaction. Of course, this must be balanced against the fact that students probably use the reserve collection several times during a semester.

The document delivery services "D" are not accorded a high impact and show substantial cost variation. However, this low impact is a reminder of the fact that we have not isolated impact with the clarity that we would like. Patrons were asked how much time "they spent using" the service". Thus, while it may take only a few minutes to request and receive a document, it may take hours to peruse that document and put it to some use.

In the middle of the pack we find two situations in which a collection of some type was assessed as a whole: "L". Finally, there were several reference services at several of the libraries, and these, represented by "R" are scattered over the whole range of the DEA plot. Thus there is clearly a great deal more to be learned about how to measure the impact, and about whether (and if so, why) the cost of achieving an impact varies so widely from one service to another.

CONCLUSIONS

The present study has shown that it is definitely possible to develop a uniform instrument suitable for use at a variety of institutions to study a variety of services. We have also shown that it is possible to develop reasonably accurate unit cost estimates for diverse functions, in a way that does not require the library to develop a complete program budget. Finally, we have demonstrated how the techniques of Data Envelopment Analysis can, in principle, be applied to the study of the relation between library functional costs and the impact of the library services or functions. Since the study on which this report is based was completed, that work has been continued (Shim & Kantor, 1998, 1999).

Acknowledgements

This project would not have been possible without several colleagues in the project. Joann D'Esposito-Wachtmann, Project Manager, supervised as many as 18 simultaneous interview studies in four geographic locations, and played a large role in the definition and testing of the successive instruments. She authored the second and third parts of the project (d'Esposito-Wachtman, 1994a, 1994b). Ph.D. students at SCILS played key roles: Michael Wilk, developed the Empirical Taxonomy, Dong Li, was responsible for data management, developing the data dictionary, and performing many preliminary analyses, Allan Zhao developed many auxiliary programs to support the analysis of transcribed texts, and Wonsik Shim did essential work in the preparation of the final report on the statistical data. The project would not have been possible without the generous cooperation of the five libraries which were the subject of the study. Our thanks to Elaine Sloan and Carol Mandel at Columbia; Frank Polach, Marianne Gaunt, Margie Epple and Emily Fabiano at Rutgers; Carlton Rochell, Nancy Kranich, Lucinda Covert-Vail, Nancy Friedland and Marlayna Gates at NYU; Donald Koepp, Nancy Klath, Mary Chaikin, Patty Gaspari-Bridges, David Goodman, and Janice Powell at Princeton; Joanne Harrar and Danuta Nitecki at Maryland, and Prof. Eileen Abels at the University of Maryland School of Library and Information Science. Financial support was provided by the Council on Library Resources through a grant to the Alexandria Project Laboratory. Mr. Li and Mr. Wilk were supported as Graduate Research Assistants during their work on this study.

REFERENCES

- Ahn, T., Charnes, A., and Cooper, W. W. (1988) Using Data Envelopment Analysis to Measure the Efficiency of Not-For-Profit Organizations: A Critical Evaluation-Comment. Managerial and Decision Economics, 9,(3), 251-253.
- Banker, R. D., Charnes, A., and Cooper, W. W. (1984) Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. Management Science, 30(9), 1078-1092.
- Charnes, A., Cooper, W. W., and Rhodes, E. (1978) Measuring the Efficiency of Decision Making Units. European Journal of Operational Research, 2(6), 429-444.
- Kantor, Paul B. (1986) Three Studies of Cost and Services at Academic Libraries. Advances in Library Administration and Management, 5, 221-285.
- Kantor, Paul B. (1989). Library Cost Analysis. In Problem Solving in Libraries, Library Trends, 38(2), 171-188.
- Kantor Paul B, Saracevic T. (1994) Studying the Cost and Value of Library Services Final Technical Report, Part 1: Models, Methods, Results. APLab/94-3/1, 82pp.
- D'Esposito-Wachtmann Joanne, Kantor Paul B, Saracevic Tefko. (1994a) Cost and Beneficial Impacts of Library Services, Part 2: Detailed History of Project Management and Processes. APLab/94-3/2, 53pp.
- D'Esposito-Wachtmann Joanne, Kantor Paul B, Saracevic Tefko. (1994a) Studying the Cost and Value of Library Services, Part 3: A Manual for Replication of These Studies. APLab/94-3/3, 21pp.
- Saracevic, T. & Kantor, Paul B. (1997). Studying the Value of Library and Information Services. Part I. Establishing a theoretical framework and Part II. Methodology and Taxonomy. Journal of the American Society for Information Science 48, 527-542 and 543-563..
- Seiford, L. M., (1990). A Bibliography of Data Envelopment Analysis. Technical Report. Department of Industrial Engineering and Operations Research, University of Massachusetts at Amherst. (1978-1990).
- Shim, Wonsik and Kantor, Paul B. (1998). A novel economic approach to the evaluation of academic libraries. Proceedings of the Annual Meeting of the American Society for Information Science, 35, 400-410.
- Shim, Wonsik and Kantor, Paul B. (1999). Evaluation of digital libraries : A DEA approach. Proceedings of the Annual Meeting of the American Society for Information Science, 36.