DOCUMENT RESUME

ED 077 539 LI 004 366

AUTHOR Childers, Thomas, Ed.
TITLE Measuring the Quality of Public Library Services; A Report to Participants.
INSTITUTION Drexel Univ., Philadelphia, Pa. Graduate School of Library Science.
PUB DATE Aug 72
NOTE 26p.;(15 References); Working report June 14-15, 1972; New Carrollton, Md.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Library Education; *Library Research; *Library Services; Library Surveys; *Measurement Techniques; *Research Methodology; Workshops

ABSTRACT

The purposes of the Workshop on Measuring the Quality of Public Library Services were to explore attitudes toward service, discuss the many facets of measuring library services and get a close-up view of one particular method of evaluating services: "hidden testing." Informal proceedings of the workshop include: (1) a talk of Ernie DeProspo entitled "The Measurement Art," giving the general state of public library measurement and some new techniques that are currently being explored, (2) two talks by Tom Childers, one entitled "Measuring Quality of Information Services," discussing recent studies that have involved the hidden testing of library services, and the other entitled "Measuring Impact and 'Quality' in Maryland," summing up the workshop and offering recommendations for future activity, and (3) "A Very Short Bibliography on Library Measurement with an Emphasis on Hidden Measurement." (Author/KE)

Measuring the Quality of Public Library Services

A Report to Participants

Workshop presented
June 14 & 15, 1972
New Carrollton, Maryland

Thomas Childers, Coordinator
August 1972

Graduate School of Library Science
Drexel University
Philadelphia, Pa. 19104

FOREWORD

In these pages you will find what is most aptly called a Report of Unproceedings of the Workshop on Measuring the Quality of Public Library Services,(New Carrollton, Md., June 14-15, 1972). The "Introduction" and Ernie DeProspo's talk have been reconstructed pretty much as they were heard. "Measuring the Quality of Library Services" is altered. And the "Closing" won't be recognizable at all, for it was never delivered, we decided to take advantage of having the group together in the Workshop, and to discuss rather than lecture. The "Bibliography" you find here is larger than the one you were sent before the Workshop, we hope that it will carry you into new explorations.

So these Unproceedings are partly a report and partly an extension of the Workshop -- intended to preserve some of the spirit of inquiry that developed there and to provide some guidelines for going beyond the limits of the Workshop.

Tom Childers

PROGRAM

Wednesday, June 14, 1972

1. Prologue and Introduction. (H. Thomas Walker, Division of Library Development and Services, and Tom Childers)

2. The Measurement Art: The general state of public library measurement and some new techniques that are currently being explored. (Talk by Ernie DeProspo; large group discussion)

3. Measuring the Quality of Information Services: Recent studies that have involved the "hidden testing" of library services. (Talk by Tom Childers; large group discussion)

4. Discussion of hidden testing: purpose, general methods, limitations, etc. (Small-group discussion, led by Ernie DeProspo and Bill Summers)

5. Briefing before telephone testing. (Small-group discussion, led by Bill Summers and Tom Childers)

6. Testing, pt. I: Simple requests for "facts"; Debriefing.

Thursday, June 15, 1972

7. Briefing before Testing, pt. II.

8. Testing, pt. II; Debriefing.

9. Participants' conclusions about hidden testing: uses, limitations, practicality, etc.

10. Measurement and the Quality of Maryland's Public Library Service. (Semi-formal presentation by Tom Childers and Bill Summers)

11. Sign-off.

INTRODUCTION

by

Thomas Childers

For a long time, I never ran across anything that shook my confidence in myself as a librarian. I was taught in library school-- and I guess I really didn't want to challenge that seriously-- that a "professional" librarian was better in one way or another than a "non-professional" librarian. I was taught that if I performed certain so-called professional tasks when I got out into the field, that I would automatically serve the public well. Oh, there were accepted ways to question what was going on. We were taught to be dissatisfied with the number of people we were serving, and we were expected to be very critical of Library of Congress classification. I don't think anyone ever mentioned to me the idea of periodically testing myself or my employees on what they were doing.

The first thing that made me uneasy with Me-as-Librarian was a report by Charles Bunge. He had tested some professional librarians and some non-professional librarians in a pretty well controlled study. He was trying to find out if the professionals were any faster than the non-professionals, and if they were any more accurate. I am not sure why I was reading the thing in the first place. But I remember my reaction to the conclusion: librarians are no more accurate in providing reference service than non-professionals; they are a little faster.

This shook me. It was probably the beginning of my skepticism.

Since then, I have done some research of my own which has upset my complacency even more; I will tell you about that in more detail later. The Bunge study and some few others, including my own, have begun to change my attitudes toward what service is and what service should be. One purpose of the Workshop is to explore our attitudes toward service.

Of course, since my own attitude would probably not have changed much without that Bunge study and without some subsequent other unsettling studies, I feel very strongly the need for systematic measurement in libraries. The purpose of this Workshop is to get us to sit down together to talk about our needs for measurement-- if there are any-- and some of the ways that we might go about measuring services. It behooves us to do so at this time, not only for our own professional curiosity and our professional commitment towards quality service but also to respond to one of the growing trends in management: PPBS. Its growth (and I assuming here that this growth will continue) necessitates the development of quantitative measures that go beyond merely counting transactions in our libraries, and begin to say something about their <u>impact</u>.

Another purpose of the Workshop is a very personal one. I want to communicate with you. And I want you to communicate with me. There is considerable isolation of formal library education from library practice. At this point in my own evolution, I'm not at all interested in one controlling the other. I am interested in sharing our knowledge and experiences toward <u>constructive</u> ends. I don't think any of us needs another academic/practice confrontation filled with the ordinary polemics and diatribes.

I have found (surprise!) that the academics do have some things to contribute. I have sort of pretested this or part of this Workshop in a couple of places where large and semi-large groups of honest-to-God practicing librarians were assembled. I have remarked at the incredulous reception of some of the research findings. The librarians were outraged, embarrassed, non-plussed, et al. To an academic, it's sort of nice to have a strong reaction-- even a violent negative one. All these goals and feelings led me to the doorsteps of the state of Maryland. And to make a long story short, the State bought the idea and LSCA is paying for it, and Drexel-- that's me-- is doing it. And, after approximately one year of planning (incredible?) here we are.

You're going to try to do at least two different things in our two days here. First, we want to talk about the general business of measuring library services. How do we measure our services now? Why? Are there better ways of measuring services? What looms on the horizon?

Second-- and this is going to be the action part of the Workshop-- we'll get our hands on a particular method of evaluation that has been tried in the last few years. In this second "action" part, it's important that we know what the specific goal is: that is, we are not here to evaluate library services in this or any other area. We are not here to draw conclusions about the level of library services. We are not here to make it hot for the folks in the State. But we are here to talk with each other about measurements in general, and about a specific kind of measurement. We are here to explore that specific method, to see how far we can push it, to find out how much we can learn from it, to find out what it can do for us as librarians in making us better on the job, to find out what kind of management information could be gotten from this kind of evaluation.

THE MEASUREMENT ART

by

Ernest DeProspo

In the public sector generally, and the field of librarianship specifically, the term "measurement" tends to generate a variety of emotional images-- for the most part negative ones. For some the term is viewed in a narrow context with connotations of highly sophisticated mathematical formulas. For others its meaning is so imprecise as to be useless. The point is, however, that for far too many librarians measurement is typically seen as a threat to both their personal and professional endeavors. The question of measurement, then, must first of all be discussed from this emotional view before we can proceed to any rational discussion of its merits and limitations.

In a teaching situation it is often popular in dealing with any particular subject to ask whether or not the subject is "an art or a science." In the area of measuring performance of library activities the response is a simple one: It certainly isn't a science, and we are not sure to what extent it even approaches an art. Part of the dilemma for the librarian is the historical fact that from the 19th Century on, we have assumed a habit towards measurement in evaluating library programs. This habit exists in the form of presenting simple statistics on either the local, state, regional or national level which imply a direct connection to performance. And part of the dilemma we face in overcoming some of the emotional issues inherent in measuring performance relates directly to the historical precedent of having established a procedure (collecting and reporting statistics) which we have come to reject as a meaningless exercise.

There are three main elements to consider within the question of measurement. First and most important, measurement assumes "evaluation." In effect, it would be inconsistent to speak of "measuring" this or that service, or this or that program, or this or that performance, and not to assume that an "evaluation" is being made of this or that service, program or performance. Second, "evaluation" assumes the desire to know where we are at a particular point in time. And third, there is the recognition that some change is needed and wanted. We should point out, however, that a key issue is that the recognition of and desire for change need not arise from the in-group. The need for change or demands for change need not come from the librarian. I think you will agree, however, that it is obviously preferable if the recognition and the initiation of change comes from the profession itself. The point is that if we can overcome or at least better understand some of the emotional issues, we can proceed on a course which would clearly be within our own interests.

Measurement, then, basically deals with these three questions. where are we, where should we be, and what are the kinds of things needed to get where we should be. Improvement in performance, improvement in programs, improvement in services, assume that we can confidently estimate levels of performance at a given point in time. Then the <u>professional judgement</u> may be made: it is either good enough, or it is not good enough and certain steps will be required to change it. Implicit in this intellectual approach is the realization that

every change must dwell within the realm of scarcity and that there are to be losses and gains as a result of the process.

What is involved is really asking the essential question: What the devil is going on here? The question suggests that there are limitations to our common sense, that a distinction between what we know and what we think we know must be made. There is verbal evidence which suggests that there are some serious weaknesses between how the individual providing the service thinks things are going and what in fact is going on. One need not go over the extensive examples of the considerable gap between, for example, the librarian's view of who is using the service and who, in fact, is using the service. Again the point is that common sense will only take us so far in answering the question of what is going on here. where we should be. and what changes are needed.

Typically, when we discuss the measurement of library services, we can anticipate a set response. For example, in the classroom the student's initial reaction is often one of bewilderment: Why do I have to study this stuff? I just want to be a librarian! That response typifies the concept of librarianship in its most narrow sense, bound by the walls of the library itself. Simply stated, it is an attitude that is inward-oriented. It tends to reinforce the observation made by Pierce Butler in 1933 when he characterized that kind of librarian as an individual happy in the "simplicity of his pragmatism." The point is that-- particularly today and for the foreseeable future-- "simplicity of one's pragmatism" is equivalent to a utopian view. That is to say, it is fundamentally non-pragmatic. Within the context of this workshop, the fact that you are here suggests that you haven't foreclosed on the possibility of looking at what you have been doing somewhat differently; that you are open to some suggestions for evaluating services (in this case, reference service) perhaps differently from the way in which you have been evaluating them; that the issue of measurement remains open; and that fundamentally you are interested in looking at this whole issue of evaluation as a way of improving the performance of libraries as they attempt to deal with the multiple demands and needs of its clientele. It is within this context that I think we need to look at some of the effective barriers to evaluation and to pull these out in the open.

Effective Barriers to Evaluation

There are four or five barriers that warrant particular identification. Jerry Walker, in a paper entitled "What's Wrong (Right) with Evaluation" (in Anna M. Gorman, editor, Proceedings of the National Research Conference on Consumer and Homemaking Education. Columbus, Ohio: The Center for Vocational and Technical Education, Ohio State University, 1970.) suggests a variety of issues which influence the process of evaluation, including some of the basic barriers to effective evaluation. Although placed in a different setting, I believe that a number of points raised by Walker are appropriate for our consideration of measurement.

THE FIRST and the most fundamental barrier is that evaluation is a threat. Within librarianship the feeling of a threat is exacerbated because we tend to "personalize" the things we do. That is, we tend to be subjectively oriented and, as a consequence, are afraid of criticism. There is a general negativism that underlies critiques of what we are doing in our professional activity, and I think it is exacerbated because we have not developed (from an educational point of view) a more objective base from which to look at the process.

If evaluation is to be considered seriously, it must necessarily embody the concept of accountability. Evaluation is a way of assessing the extent to which any program, project, activity, service, etc. meets its own specifications, its own objectives. And in order to see if these specifications are being met, they first must be stated in clear and demonstrable terms. For example, within the area of reference we would say that a demonstrable objective is to satisfy 75% of all the reference requests; in circulation, we could say that the demonstrable objective is to satisfy 80% of all circulation demands; and so forth.

Historically, public agencies (and libraries are no exception) have been very hesitant to state their objectives in terms as precise as those. The fact is that we don't really like to set forth our intentions in measurable terms. And, until recently at least, it has probably been functional not to do so. Today, with the increasing strains on fiscal resources, with increasing pressures to open up the decision-making process as libraries are forced to relate more to other institutions, we may see some significant changes. For example, the pressure to move into PPBS, and the very poor experience which we have had with it so far, relates in part to this issue. That is, we really don't have a data base upon which to comply with the PBB requirements. We don't have the experience in generating the data needed for matching against the objectives of that kind of budget.

In other words, measurement and evaluation demand that we do just what we have historically avoided doing. And, when you couple this situation with the emotional issues involved in measurement and evaluation, you can see that change, if it is going to occur, will probably have to take place in stages. We will need to develop appropriate strategies and tactics to decide which areas, which activities have top priority for gaining the kind of experience which will allow us to set forth procedures in terms that we can then translate into measurement indicators. Again, the point is that we need to know where we are first before we can proceed to the next step on a logical basis. Typically, needing to know where we are first is often the less interesting question. We would much rather know the meaning of this or that, and there is tremendous pressure to skip steps 1, 2, and 3 and jump into 4, 5, and 6. I firmly believe that skipping these basic steps has, in the long run, been self-defeating. And let me note that we probably can be much less ambitious than is commonly assumed in deciding what kind of basic information is needed before proceeding to some of the more meaty issues that face us.

It is important to point out once more that both as individuals as well as professionals, very few of us want to be evaluated. Teachers don't believe that they really can be evaluated effectively as teachers; students reject the notion that grading is fair or appropriate. We have to face openly and honestly the reality that evaluation is considered a threat if we are going to overcome one of these effective barriers. Unless we do so, any rational attempt to evaluate what we are doing in libraries is likely to fail.

A SECOND obstacle to evaluation is that the library decision maker or policy maker can <u>respond</u> to situations much better than he or she can <u>plan</u> <u>and</u> <u>create</u> a situation. The routine pattern of any activity (again libraries being no exception) tends to create actions which are really programmed. The problem with change is that it deviates from the norm. It tends to heighten the insecurity

and the doubt because we can't resort to a programmed response. We have to come up with a non-programmed response. And this is always more difficult, more uncertain, more threatening. We typically cannot specify information needs completely and accurately in advance, or specify the types, the priorities, and the weights of the criteria which we use in making our decisions. Yet on a post-facto basis, we can tell the evaluator what we did not want and what was missing.

A THIRD barrier is that an evaluation serves multiple purposes and these purposes often contradict one another. For example, within a particular program or service, you could think of a supervisor who looks at the purpose as really kind of superficial, the evaluation as really being done to appease some funding agent. The evaluator may view the purpose as a careful reporting of the information required to make better decisions. A client of the service may view the purpose as unimportant and just another waste of tax dollars. These conflicting attitudes are common in any evaluation process, and because of them the evaluation tends to generate far more questions than it settles.

THE FOURTH barrier to evaluation is money; it is an expensive proposition in terms of time, personnel resources, and money. And in time of scarce resources evaluation is often seen as a luxury which can be sloughed off when the belt needs tightening.

Finally, A FIFTH and important barrier to evaluation is that it is often seen as a one-shot affair. Rather than an on-going operation, it happens once. Therefore, the time, effort, and emotion involved are not really worth the effort. It is not going to have any <u>continuing</u> influence or effect on the program.

No amount of rhetoric will <u>eliminate</u> the existence of these barriers to evaluation and measurement. The first order of business is their recognition. With respect to evaluation as a threat, I think we have to accept this reality as one of our human frailties. But we also must recognize that if evaluation is initiated and controlled wisely by the profession, rather than being conducted by some unknown "outsider", it becomes that much less threatening. Again, I think we must realize our need to reduce the subjectivity with which we view what we do and understand that some objective, systematic evaluation may very well be in our own best interest.

The observation that the policymakers can better <u>react</u> to situations than <u>anticipate</u> them basically reflects the reality that our hind-sight is always clearer than our foresight. It is a reality-- certainly not unique to libraries-- which requires some change in administrative style. It will require policymakers to be more willing than they have been to go out on a limb occasionally-- even if it cracks. Easier said than done. But we need somehow to find better ways of moving into non-programmed decision making and to recognize this effort will require trial and error.

The facts that evaluation serves multiple, conflicting purposes and is expensive, are both inevitable. Again, recognition of these factors will reduce their impact. It may be that with respect to costs we have been penny-wise and dollar-foolish. Better knowledge of what we are doing, why we are doing it and how to improve what we are doing may be an investment that will, if not produce dollar savings, at least produce substantial improvements in our products.

With respect to the claim that evaluation is a unique or one-time affair, the simple response is that it need not be so and that it is less likely to be so the more the individuals in the process are involved. In our case, the more the librarians in a particular library are involved in evaluating what they are doing and in interpreting the results for that particular library, the less likely is the activity to be a one-time thing.

The Art of Measurement within Librarianship

I indicated earlier that we have a precedent for measurement dating back to at least 1887 with the adoption of a policy by the American Library Association to collect statistics for "measuring" what libraries are doing. This historical precedent has become a substantial liability in moving to other forms of evaluating and measuring performance. The traditional approach-- counting things and reporting these things in some kind of table or form-- has generated a process of looking at what we do from an inward rather than from an outward point of view. Simply stated, we tend to be "thing" or input-oriented; we assume that these inputs equal service. For example, we assume that the number of books that we add (which is an input) equals or has something to do with the quality and quantity of service (which are outputs); we assume that circulation is an indicator of services; we assume that the number of professionals is an indicator of service; and so forth. The point is that most of our measures that we still heavily rely on are not user-oriented. We are forced to make the fallacious jump from input to output. Yet, as we are challenged by others (such as budget agents) to prove our worth, we become less and less confident in the relationship between the inputs and the services which we offer.

The basic controversy with respect to measurement in the field of librarianship can very easily be summarized. The position is essentially emotional, and it is the most significant argument made. Those who deny that we can measure in any quantitative way what we are doing argue that you can't infer quality from quantity. The problem with that argument is that, historically, we have been inferring quality from quantity in the kinds of statistics that we report. On a logical basis we know that (1) there are relationships between quantity and quality and (2) absolute precision is not only unattainable, but not necessary. Our space program tolerates a margin of error, even though this endeavor is seen as "absolutely precise." The issue is not whether one can infer quality from quantity, but what are the professional judgements and interpretations required to make what type of inferences. Our overall lack of experience and the growing challenge to the manner in which those in the profession have been defending their programs has made us rather timid in saying what, in fact, quantitative measures mean. In large part, then, we are dealing in a situation in which we simply haven't generated sufficient experience in looking at different ways of evaluating performance.

Within our American Library Association study on the measurement of public library effectiveness we have attempted to develop measures which simulate the library user as he seeks very basic library services. Our concern is get at steps 1, 2, and 3 before we proceed to the more interesting questions that are in steps 4, 5, and 6. In our extensive literature search on measurement we found no public agency doing well. We are not atypical, then. For example, in spite of the enormous dollar outlay, there has been no method devised to

measure the effectiveness of public schools-- despite all the testing and all the claims. This fact isn't to suggest that we should be satisfied by how much we haven't been doing, but simply to indicate that we are not atypical.

We find that most of the library statistics presented are descriptive and are very limited in the kinds of analysis which can be made. The need for more sophisticated statistical approaches is obvious.

Applying a variety of statistical tests to the U.S. Office of Education statistics, we find very little relationship among the items reported. The only possible conclusion that we can arrive at is that smaller units of service appear to be more effective than larger units of service. The study has a number of constraints-- of course the obvious ones are time and money. But more importantly, the decision was made that the methods and approaches developed would be influenced by the fact that the data should be collectable at the local unit and that final interpretations should be made at the local unit. It is one thing for a highly sophisticated trained team of researchers to go out and collect information, come back, manipulate the information and come up with their findings. It's quite another thing for those who weren't trained in the process to do so. Yet the former approach is one which typically tends to be a "one-shot" operation. You seldom see the researcher again-- or more importantly, see and understand the results of the research.

So considerable effort has been placed in developing an approach which would involve librarians in all phases of the process of gathering the data and then in final interpretation of the data. For example, we put on a series on two-day workshops in which librarians from each of the libraries involved in this study was given instructions on how to supervise the collection of information by their local library. Rutgers has been active as sort of a central data processing location from which preliminary data analysis will take place. From there, information will be sent back to the local library for final interpretation by the librarian.

The study has two distinct phases, broken down into data collection over a two week period. In the first week, information is collected on the following kinds of items: library sponsored and community oriented programs; document availability (for document availability we essentially want to know what chance the user has in getting a book or a journal article from that library); the physical plant and equipment; staff, including library experience and education.

The second week-- which involves three days in which we ask each library to indicate their high-, medium-, and low-use days-- is set up to collect data on the following kinds of items: how the facilities and equipment are used; what the patterns of reference usage are; what outside library circulation looks like; and what library use looks like. Every user entering the library over these three day periods (the high-, medium-, and low-use days) is provided with a very small library user "ticket" which gives us some information about who is using it-- student, non-student, sex, occupation-- and also gives us information on how long they used the library so that we can say something about that library in terms of time usage.

We have found the tickets to be particularly discriminating. For example, we were able to characterize our four pilot libraries as: (1) a "supermarket" operation, in which the average turnover is about 15 to 20 minutes; (2) low-turnover, older user group, with the average user in the library close to an hour; (3) program-oriented, with highly competitive divisions in the library. Each library tends to have unique patterns which might have been used to predict such patterns; yet standard statistics or measures did not suggest that the patterns existed.

Overall, we find a very posit tiude toward the library by the user. We see very poor utilization of skill resources (of professionals, in particular) in the libraries. In the pilot libraries we find a very low level of reference activity. The reference questions, by phone or in person, average something like ten an hour. Fifty percent of these questions are directional in nature.

We are just beginning to analyze the data beyond the pilot libraries, so all these comments are highly preliminary in nature. But let me make a final point: The high degree of involvement of each librarian has produced rather remarkable enthusiasm. The study reinforced my own biases when we talk about measurement and evaluation. If we want evaluation to be an on-going process, then it is critical that the librarian be involved and understand what is going on. For this reason, I strongly support this kind of Workshop and highly commend the efforts of the state of Maryland and, specifically, Dr. Childers.

And a footnote: The matters of change and adaptation are directly related to evaluation and measurement. I think we need to take a more than casual look at them. Let me conclude with a statement from Alvin Toffler's Future Shock:

> Our psychologists and politicians alike are puzzled by the seemingly irrational resistance to change exhibited by certain individuals and groups. The corporation head who wants to reorganize a department, the educator who wants to introduce a new teaching method, the mayor who wants to achieve peaceful integration of the races in his city--all, at one time or another, face this blind resistance. Yet we know little about its sources. By the same token, why do some men hunger, even rage for change, doing all in their power to create it, while others flee from it? I not only found no ready answers to such questions, but discovered that we lack even an adequate theory of adaptation, without which it is extremely unlikely that we will ever find the answers. (Bantam edition, p. 3)

MEASURING THE QUALITY OF INFORMATION SERVICE

by

Thomas Childers

The History

The measures of library "service" in the past and today are nothing much more than simply counting. We count the numbers of documents circulated, books added, dollars spent, hours open, square feet, chairs, and on and on.

For many years we have assumed that "bigger is better." We have assumed that more circulation meant better library service; that more books per capita, more dollars per capita, more users per 100,000 population, more hours open-- that these add up to better library service. It has only been within the last few years that there has been even a piddling effort to test this assumption: Does bigger mean better?

Moving a step further: What does "better" mean, and finally, How good is "better?" As we continue to measure the so-called quality of our services in terms of simple size and frequency, we seem to reinforce the notion that the quality of our services is determined (caused) directly by numbers-- that service is "improved" by getting people to take out more books, or by registering more borrowers, or by acquiring more books, or by handling more reference questions.

Yet, somehow we know that no one of these remedies-- nor perhaps all of them together-- necessarily means that we are going to improve the level of the service that we provide to the public. Somehow we know that these factors may not be critical in determining the quality of service that we offer.

Here we sit observing the things that are most observable. Which is natural. And it certainly tells us things about demands put on the system, queuing, transaction frequency, carpet wear, etc. But to make statements about the quality of our service, I think we must begin to go beyond simply tallying this or that thing that can be easily tallied. Evaluation of service quality is probably going to be more complex.

I don't mean to suggest that the library field is any worse off than almost any other field. It turns out that even some executives of profit-making corporations are having difficulty in recent years of defining just what is "profit." What may appear to be profit in the short term can very often turn out to be loss in the long term. The gross national product (GNP)-- which at one time seemed to be a very respectable measure of the quality of American life-- in the last few years has come under considerable attack.

We may all be in this thing together. Maybe net profit, GNP, and the number of books circulated in a year are none of them useful in talking about the effectiveness of our corporations, or the quality of our life, or the level of our library service.

The measurement of reference/information services is very much in the ballpark. Since the early 30's, we have tallied the numbers of questions answered and unanswered, kinds of sources used (dictionaries, almanacs, indexes, etc.), the number of sources used, the kinds of questions answered and unanswered, and so forth.

Measuring the Quality of Information Service

It seems that we never really asked ourselves how well we are providing our reference and information service.

Testing the performance of people at the reference desk began-- as far as the literature goes, anyhow-- in the late 50's with the report for the Commissioner of the New York Department of Education.[1] In trying to get at the comparative levels of services from library to library in New York State, some testers went from library to library and asked a number of questions. The librarians were supposed to answer the questions as quickly as they could within their own individual collections.

The performance of the libraries was not specifically recorded. But what is most important about this study, anyhow, is that it was done. One small part of library services were being tapped at the point where it meets-- or might meet-- the client. Hurrah! There are some problems in the design of this study that appear on the very surface: the testers were identified as testers; and the librarians were limited to what they could find within their own collection. The situation must have felt to the librarians like a Graduate Record Exam-- or at least not exactly like a bona fide encounter with a real live client.

Around 1965, Enoch Pratt Free Library tapped their own telephone reference lines over a period of four months. They concluded that about one-third of the questions were not answered satisfactorily (whatever that means). It was never broadcast through the land; probably the land wasn't ready for it at the time.

Long ago, those cream-colored plastic towers of academia, the social psychologist coined the phrase "cognitive dissonance." In English that meant something like "shock" plus "disappointment." Can you imagine the amount of cognitive dissonance-- shock plus disappointment-- that would have run through libraryland if Pratt had broadcast its results? After all, at that time-- and for the most part, still-- we like to think that we are performing about 100%.

(An aside: Last year I did a very tiny study of a city library and some suburban branches in the mid-west. The results of the testing that I did were very disappointing-- that is, shocking-- to the central library director. I am certain that he would have had my head or reputation or both if I had not mentioned Enoch Pratt's study. I was very happy to note the strong effect of such a simple statement as "Well, when Enoch Pratt studied itself, it seemed to be performing at a 66% level of success." That irate director got as calm as a kitten; in fact, he started to think that his performance wasn't so bad after all.)

Findings in the Recent Past

Terry Crowley at Rutgers in 1967 got interested in the idea of acting like a client and then judging the kind of response you get from librarians. So he got a bunch of questions together and bunch of shills, or proxies, or shoppers, and started asking those questions at libraries in Pennsylvania and New Jersey. Some of the questions they asked in person and some of them they asked over the telephone. They included all kinds of questions, but they were mostly questions whose answers had changed in the last few days or last few weeks or last few months. Crowley's interest was in finding out how well librarians kept up with the world as it is. He had a suspicion-- he really had a point that he wanted to make-- that when, say, World Almanac is put into a paper binding, a part of every librarian is bound right in there with it and can't escape. A "fact" that is anchored in World Almanac remains a "fact" until the next edition, whether it has changed in the real world or not. To some librarians, a local senator who has died is not dead till duly recorded in a library "tool." Or, a change in monetary systems in Australia does not "exist" until it is published.

Crowley's thesis was that the librarians are not on top of the world that is around them and to that extent cannot be effective disseminators of information. At the same time, Crowley was interested in finding out if libraries that are richer performed better than libraries that are poor.

Hold your hat. The results were:

First: The richer libraries did not perform significantly better or worse than the poorer libraries.

And Second!: The overall performance was 54.2%. A little better than half the answers contained the correct answers. Among the problems in answering, Crowley found that transposition of numbers and the failure to up-date the information were the most common.

While dramatic, the Crowley study did not have a lot of questions in it, nor was the sample of libraries very large. So I figured what he observed could have been attributed to chance. That is, I was not convinced that libraries perform that poorly.

Until I started doing my own studies.

I was interested in testing the performance of libraries in the way Crowley had; but I was most interested in how performance correlated with a lot of other library factors-- such as collection size, number of professionals, books, hours open, dollars spent per capita, books per capita, circulation, and on and on. I won't entertain you with all of the figures-- one single chart had 1,081 4-place statistics of correlation on it, and had to be spread over seven whole pages-- even though many of the statistical findings were sort of fun and even useful. But what is really important for us now and what-- I must confess-- shook me, and still shakes me a little-- is the performance of the libraries.

My study was quite different from Crowley's. He had a limited bunch of libraries he was asking questions from and a very small number of questions-- which is characteristic of an exploratory study. In my research, however, I hired a

gang of people to ask 26 questions at each of 25 libraries. With a few thrown out for this reason or that, the total number of encounters with libraries was over 600.

And what was funny (but then again, not so funny) was that the libraries in my study performed at the same level as in the Crowley study. Depending upon how you judge "correctness," the libraries in my study were successful between 50 and 60% of the time. Unlike Crowley's, though, I did not find much in the way of incorrect responses. Usually, either the "correct" answer was given, or no answer was given or a semi-answer was given. (A semi-answer might be something like, "Well, I think it might be _____.")

Since I did that large study in New Jersey, I have done a couple of smaller studies in two states. In one state the results were approximately of the same level, 50 - 60%; in the other state performance level was way down, around 20 - 40%.

Before you start throwing things at me or your reference staff, let me hasten to say that there were lots of limitations in these studies. The questions-- even though they were supposed to look like real library questions-- and the callers-- even though they were supposed to look like real library patrons-- we have to assume did not cover the whole range of possible library demands and library clients. It is possible that each study had a number of biases within it. To that extent, we are on shakey ground if we go around condemning public libraries for providing universally low quality service.

What can we say? Simply, WE MAY NOT BE DOING AS GOOD A JOB AS WE HAVE ALWAYS ASSUMED.

If a study that occurs inside of those ivied ivory walls (assuming that it gets outside of them) makes us a little uneasy about what we are doing, makes us question our performance, then the research may have been worth it all. This particular method of hidden testing has caused some people to worry a little over what is happening in the reference departments. It has caused a few people to shout at me and get angry at themselves, because the data that it produced is so foreign to them. It is not surprising; it's like going through a kind of culture shock.

The "Hidden" Technique

This is coming out backwards. I told you the results of some studies; now let me go into how we go about them. As I said before, the method really consists of asking a bunch of questions at a bunch of libraries in such a way that both the people who asked the questions and the questions themselves appear to be bona fide.

Twenty-five libraries were asked each of 26 questions, by the 26 proxies. The questions were for the most part fairly simple. Some examples:

> What is the gap specification for a 1963 Chevy Corvair?
>
> Could you identify this piece of poetry for me, please:
> "O pardon me, though bleeding piece of earth..."?

What was the federal budget expenditure in 1936?

How many families are there in (this) county?

What does the O.J. in O.J. Simpson stand for?

What is the book that the Peter Sellers movie, "The Battle of the Sexes," was taken from?

What were the exact words of John F. Kennedy when he was asked how he became a war hero?

One of the major problems in hidden testing is to "control" the application of the question. This means that, as much as possible, you not only want a bona fide question and a bona fide caller, but you also want the application of a question to be consistent from library to library. That is, you don't want a pleading, cajoling tone used with one library, and a business-like, straightforward tone used with another, for the same question. So to help control for this, we have one proxie call the same question to every library in the sample. By no means do you get absolute control in this way, but it is probably safe to assume that you will get more consistent application of a question this way.

The caller is also given a specified amount of information to include in each question. If the question is: "Is the Australian money system a decimal system?" he gives nothing more and nothing less. If the question is: "How far is Athens, Georgia, from here?" nothing is left out and nothing is added. He does not omit any information, nor does he include any extra information. So, when the respondent hesitates and says "My goodness. I just don't know..." the caller doesn't say, for instance, "I think the Australian money system has changed the last few years." He says something neutral and information free like, "I'd like to know" or " ."

This may seem artificial. And it is. But this is the way we tried to control the way the libraries were stimulated, in order to be able to compare the performance of one library to the performance of another.

Since we were interested mostly in how correct the libraries' answers were, and not in what kinds of questions they would accept, or from what kind of patron, we developed a list of questions that looked "legitimate", and assembled a core of proxies that looked and sounded real, too. None of the proxies (well, we made a mistake on one) sounded like a school student; and if there were a chance that a question might be taken as a request for information in working out a crossword puzzle or a homework assignment, we clothed everything in a "personal rationale" that made it sound like a "legitimate" adult request (For instance, "My father needs this information in order to fix his car."). Likewise, in constructing this rationale, we were careful to prepare ourselves so that the respondents couldn't expect any further information out of us. For example: "Gee, I don't know. It's for my boss."

To keep from biasing the studies against times when professional librarians would not be there, we started at random times during the day and week in our calling. And in order to give the libraries the benefit of the doubt, we allowed them to take as long as they wanted in answering the question, and to

call back-- or rather to be called back. Calling back posed somewhat of a limitation. We could not allow the library to call us back. In almost every instance, even when the request was for a mere simple fact, they wanted to call us back. But since we were calling many libraries long distance, if we had allowed them to return our call, we would have been given away as non-residents. So, to avoid suspicion, we usually made an excuse as to why we would have to call them back. ("Gee, the baby's asleep." "I'm calling from work, and really shouldn't have incoming calls.")

The proxie writes down as much of the whole encounter as he can, verbatim. Some people can do this quite easily; and it's very difficult for others (me, for instance).

Responses are then judged. For judging the responses it seemed best for me to have ready before the question was asked a "criterion of correctness." For example, you might accept something as simple as "Shakespeare's Caesar" which could be predetermined as correct instead of the longer "William Shakespeare's Julius Caesar, act III, scene i." This may seem an awfully fine point, but immediately when we got the answers back it was apparent that to be fair from library to library, and to be consistent in judging them, we had to have a concrete standard such as the "criterion." We found it worthwhile, even so, to have several people judge the same responses.

What can we learn from these efforts? Up to now, the major interest has been in looking at the correctness of the answers supplied by the libraries. But there are other facets: speed, attitude toward various 'types' of callers, willingness or readiness to refer, consistency of policy and service, etc., etc. Look even beyond that, and you will possibly see other things that can be learned from "unobtruding."

This means that we have to make some very personal assumptions about what "quality" service means: does it mean speed, accuracy, pleasantness, familiarity with materials, skill in interpretation, general knowledge...? As much as anything, we are probably testing ourselves, our own concepts of what library service is or isn't and what we can expect from places called libraries.

Remember that what you uncover in your digs may be ugly. Of course, there are people like the woman who knows the spark plug gap specification for a '63 Chevy Corvair off the top of her head; she turns out to be an amateur mechanic.

But then there'll be the respondent who, encountering a statistic that reads "$8,493, in millions," as the federal budget expenditure for 1936, reads it over the phone as "8 million, 4 hundred 93 thousand dollars."

And the sweet young (judging from her voice) thing who found the page in Bartlett's with the quotation "O pardon me, thou bleeding piece of earth..." and correctly identified it as coming from a work of Shakespeare.

"But what's it in?"

"Well, I dunno...er..."

"Is it part of a play or poem or something?"

"Uh, well, I can't tell. For other quotations it says the play. Here's one...one from Macbeth, and Julius Caesar. And Twelfth Night. An then there's a lot from Ibid."

"Ibid?"

"Yeah. I-B-I-D."

"Well, thank you."

"You're welcome."

FOOTNOTE

[1] Report of the Commissioner of Education's Committee on Public Library Service. Albany: The University of the State of New York, The State Education Department, 1958.

CLOSING: MEASURING IMPACT AND "QUALITY" IN MARYLAND

by

Thomas Childers

Just like most other active states (as opposed to dormant ones), Maryland has been concerned with the level of service it provides. Just like most other states, too, Maryland has been primarily concerned about quality level, or performance, at a point that is removed from the point where the user meets the library services.

Take, for instance, the Statistics of Maryland Public Libraries: 1969-70. Fourteen "measures" that say something about Maryland's libraries are recorded here. Only one (count 'em, one) treats of the user and his use of the library. The rest say things about inputs: books, dollars, staff, etc. While we have to deal with service inputs, of course, that isn't all there is to knowing what we're doing or what we are, by a long shot.

We assume that all kinds of things ultimately have an effect on the kind of service we give. This assumption is reflected in some of the studies done for the libraries of this state:

> Feasibility of joint service centers for metropolitan libraries, by Duchac in 1968.[1]
>
> Study of library aids, or associates, in Maryland, by Clark and Johnson in 1971.[2]
>
> Proposal for a legislative package by the Commission on Public Library Laws, 1970.[3]

These efforts and their recommendations-- and I'm not questioning the need for them, or their usefulness-- are presumed to have an effect on library service or they wouldn't have been undertaken. However, the recommendations are so far removed from points of "impact" on the actual library user that it's hard to predict how he might benefit from them.

There is a group of studies in the State that gets closer to the user, though.

> Students and the Pratt Library.[4] Talks about frequency of use, how use fits into personal daily life, what the library is used for, and like that.
>
> Baltimore Reaches Out.[5] Deals with reading; media use; relative use of libraries as suppliers of materials, compared with other sources of materials (which comes close to measuring the impact of library service in general).
>
> Bundy's metro-Maryland study.[6] Primarily deals with the habits of users. Says little about the success of service. You might infer that the libraries perform well for the few people who are very heavy users, and that they perform poorly for non-users. This doesn't tell us much about performance, though.

Warner's study of ILL requests.[7] Reports 60% success; implies a level of service from the user's point of view.

We come even closer in the Consad Report: Baltimore Metropolitan Area Work Program, 1970.[8] The report recommends a systems approach to planning for the development of libraries in the area-- which includes

1. development of concrete goals

2. creation of measures of success, and the mechanism for measurement

3. actually measuring 'success'

If specifically recommends the creation of a Planning Information System. Included as part of this System would be (a) regularly collected data and (b) intermittently collected data (probably for the figures that are hard to collect all the time). The report echoes other sources in saying that PPBS points up the need for more and better data. But most important, the data SHOULD BE MORE THAN INPUT DATA, more than books, staff, hours, dollars; and it should go beyond the very rudimentary output data that we now have-- which, statewide, for Maryland is essentially "number of books circulated." There is a need for more output data, in more detail, with more meaning (which means, probably, closer to the point where the client consumes the service).

What we dealt with for two days in the Workshop was just one humble possible technique for looking at outputs. You have been introduced to a technique, a tool for internal management and possible for statewide development. You may or may not have been impressed once again with the importance of looking critically at what libraries are all about, from the client's point of view. Maybe some of the myths of service have been-- if not dispelled-- at least unsettled a little bit.

Don't stop there. Let me recommend that you (a) communicate with your coworkers back home, and with other libraries in the vicinity. Communicate anything you want, communicate the Workshop. You'll each have your individual message, and it will probably relate to the need to look at service outputs. Proceed within your libraries, branches, systems, to question in a heretical, sacrilegious way the assumption that libraries are performing 100%.

Proceed carefully, though. Remember the efficiency expert in the movies of the forties and fifties? Why was he always cast as a son of a bitch? Because he was. Not due to his own fault, but due to the managers who hired him and who weren't careful in using the tools of evaluation that he developed for them. The managers in those olden days tended to use measurement ON the people in the organization, rather than FOR them. Management is not the authoritarian game it once was, of course. But measurement can still be abused. That old efficiency game (systems analysis, operations research, PPBS) can threaten people as well as help them do a better job. Caveat.

(b) Start developing your own devices for measurement, things that will help tell you what you look like-- especially from the user's point of view. The method we have called hidden testing is just one small possible tool.

(c) Consider working more closely with the University of Maryland Library School and the State Library toward a statewide effort to

1. experiment with impact and 'quality' measures, and

. collect regularly certain impact and quality data for statewide consumption.

FOOTNOTES

[1]Dwinac, Kenneth F. A Library Service Center for Suburban Maryland County Systems. Baltimore, 1968.

[2]Clark, Philip M. and Johnson, Mildred Y. A Training Program for Library Associates in Maryland. New Brunswick: Bureau of Library & Information Science Research, Graduate School of Library Service, Rutgers University, 1971.

[3]Commission on the Public Library Laws. Libraries for the People of Maryland. 1970.

[4]Martin, Lowell A. Students and the Pratt Library: Challenge and Opportunity. Baltimore: Enoch Pratt Free Library, 1963.

[5]Martin, Lowell A. Baltimore Reaches Out: Library Service to the Disadvantaged. Baltimore: Enoch Pratt Free Library, 1967.

[6]Bundy, Mary Lee. Metropolitan Public Library Users: A Report of a Survey of Adult Library Use in the Maryland Baltimore-Washington Metropolitan Area. University of Maryland, School of Library and Information Services, 1968.

[7]Warner, Edward S. Maryland Interlibrary Loan Network Loan Requests: System and Units Analysis. Baltimore: Maryland State Department of Education, Division of Library Development and Services, 1969.

[8]CONSAD Research Corporation. Baltimore Metropolitan Area Library Study. Pittsburgh: CONSAD Research Corporation, 1970.

A VERY SHORT BIBLIOGRAPHY ON LIBRARY MEASUREMENT WITH AN EMPHASIS ON HIDDEN MEASUREMENT

(The entries are in order from the most general to the most specific. The Bunge and the Crowley-Childers items are fairly formal reports of dissertation work in the area of testing the "quality" of library service.)

Line, Maurice B. Library Surveys: an Introduction to Their Use, Planning, Procedure and Presentation. Hamden, Conn.: Archon, 1969.

In 134 pages, a respectable, straight-forward introduction to the study of libraries in more or less conventional ways. Includes design of the study, collecting the data and analyzing it, and reporting the findings.

Webb, Eugene J. and others. Unobtrusive Measure: Non-reactive Research in the Social Sciences. Chicago: Rand McNally, 1966. (paper)

As a whole, it's a sprightly introduction to research methods and well worth buying. The section entitled "Intervening Observer:" (pp. 155-164) presents hidden or contrived observation as a technique of measurement.

Drott, M. Carl. "Random Sampling: a Tool for Library Research". College and Research Libraries, 30 (March 1969): 119-125.

Questions about the accuracy of library records, the behavior or attitudes of patrons, or the conditions of the books in the collection can often be answered by a random sampling study. Use of this time and money saving technique requires no special mathematical ability or statistical background. The concept of accuracy is discussed and a table is provided to simplify the determination of an appropriate sample size. A method of selecting a sample using random numbers is shown. Three examples illustrate the application of the technique to library problems. -- abstract quoted from the journal.

Rothstein, Samuel. "The Measurement and Evaluation of Reference Service". Library Trends, 12 (January 1964): 456-472.

One of the most comprehensive treatments of the ways we have looked at reference service over the years. It amounts to a digest of traditional measurements of the reference function in public libraries.

Bibliography - continued

Childers, Thomas. "Managing the Quality of Reference/Information Service". Library Quarterly, 42 (April 1972): 212-217.

A brief review of measurement of reference/information service in libraries leads to the observation that few measures have dealt with the service product from the user's point of view. The late fifties saw the beginning of tests of reference efficiency. In the late sixties, unobtrusive (hidden) testing of the quality of service was begun. The majority of this paper is devoted to describing the method of such unobtrusive measurement. Findings are reported briefly as dramatic illustration of the need for further investigation. Finally, applications are proposed. The method can be seen as a quality control, with value both for the monitoring of service quality within a library and for establishing nationwide standards. -- abstract quoted from the journal.

Bunge, Charles A. Professional Education and Reference Efficiency. (Research Series No. 11) Springfield: Illinois State Library, 1967.

An outstanding research effort that attempts to relate the characteristics of librarians to their success at the reference desk.

Crowley, Terence and Thomas Childers. Information Service in Public Libraries: Two Studies. Metuchen, N. J.: Scarecrow Press, 1971.

Two of the first unobtrusive investigations of the question-answering capacities of public libraries. Libraries covering a range of annual expenditures from 10,000 to 700,000 dollars were tested on their responses to 36 different questions.

LIST OF PARTICIPANTS

Ruth Almeida (Mrs.), Branch Librarian
North County Branch
Annapolis & Anne Arundel Co. Pub. Libr.

Carolyn Andrion, Reference Librarian
Harford Co. Pub. Libr.

Marie Barrett, Senior Library Assist.
Calvert Co. Public Library

Charles Blank, Assistant Director
Washington County Free Library

Howard Borthwick, Deputy Director
Montgomery Co. Public Library

Scott Bunn, Adult Asst., Gen. Ref. Dept.
Enoch Pratt Free Library

John Burgan, Chief
Extension Division
Enoch Pratt Free Library

Jan Callison, Asst. Branch Librarian
Severna Park Branch
Anne Arundel Co. Public Library

Hugh Courtney, Branch Librarian
Bowie Area Branch
Prince Georges Co. Memorial Library

Anne Dalzell (Mrs.), Asst. Chief, Pub. Serv.
Montgomery County Public Library

Betty Day (Mrs.) Reference Librarian
Montgomery County Public Library

Irma Dillon, (Mrs.), Asst. Coord. for Young Adult Services
Baltimore County Public Library

Lance Finney, Specialist, Library Facil.
Division of Lib. Devl. & Services
Maryland State Dept. of Education

Lois Frost, Bookmobile Supervisor
Annapolis Area Branch
Annapolis & Anne Arundel Co. Pub. Libr.

Judy Getz, Senior Children's Spec.
Cockeysville Branch
Baltimore Co. Public Library

Mary Hall, Asst. Director for Pub. Serv.
Prince Georges Co. Memorial Library

Patricia Harris, Extension Librarian
Annapolis & Anne Arundel Co. Library

Katherine Hurrey, Supr., Patron Services
Southern Maryland Reg. Libr. Assoc.

Joyce Jennings, (Mrs.), Young Adult Coord.
Annapolis & Anne Arundel Co. Pub. Libr.

Adele Jones, Principal Libr., Reference
Oxon Hill Branch
Prince Georges Co. Memorial Libr.

Nancy Liss, Supr. Information Specialist
North Point Area Branch
Baltimore Co. Public Library

Nancy Maier, Personnel Officer
Baltimore County Public Library

Nancy Mathers, Reference Librarian
Hyattsville Branch
Prince Georges Co. Memorial Library

Donald Napoli, Asst. Branch Librarian
Loch Raven Branch
Baltimore Co. Public Library

Geneva Robinson, Coord. of Adult Services
Baltimore County Public Library

Betsy Shorb (Mrs.), Young Adult Librarian
New Carrollton Branch
Prince Georges County Memorial Library

Fern Solomon (Mrs.), Reference Librarian
Montgomery County Public Library

Anne Sweat, Asst. Coord. of Adult Services
Prince Georges Co. Memorial Library

Participants - continued

H. Thomas Walker
Division of Library Development & Services
Maryland State Dept. of Education

Eleanor Wasielewski (Mrs.), Asst. Dir.
Harford County Public Library

Barbara Watson (Mrs.)
Howard County Library

WORKSHOP STAFF

Thomas Childers, Coordinator
Graduate School of Library Science

F. William Summers, Assistant Coordinator
College of Librarianship
University of South Carolina

Ernest DeProspo, Speaker
Graduate School of Library Service
Rutgers University

Barbara Dunn, Graduate Assistant
Graduate School of Library Science
Drexel University