Theory Into Practice

Patricia M. Kelley

Performance Measures: The Theory

Rightfully so, an academic librarian who is considering the implementation of performance measures will question whether the usefulness justifies the time and effort required. After all, don't we already know whether or not our libraries are doing the best job possible with the resources at our disposal? The short experience with performance measures of Gelman Library at The George Washington University indicates that the time and effort are well spent and that measures help to provide objective evidence to support or refute our intuitive professional evaluations of how well we are serving our community.

The data we have gathered offer few surprises. Like librarians everywhere, we have a fairly good sense of where our successes and difficulties lie. Our dilemma is that each person is familiar with a few pieces of a puzzle that portrays a complex service organization. The shapes of our puzzle pieces change continually, however, with the introduction of new technologies, the rise and fall of budget allocations, turnover of staff, programmatic changes in our parent institution, and resource sharing opportunities. As we work with each other and with the faculty and administrators outside the library to ensure that the puzzle pieces continue to fit together properly and that the picture they form is pleasing to this particular university, we find that we need to describe library operations in concrete terms. We need to describe objectively the state of the library to ensure clarity of communication and to give credibility to the assessment we make about how well the library is serving students and faculty. Performance measures provide that description. They can be used to explain what the library is achieving and what resources it needs. When compared with standards, they describe how well the library is performing. And when compared with organizational goals, they tell us how well we are serving our target clientele.

Patricia M. Kelley is Assistant University Librarian for Programs and Services at The George Washington University in Washington, D.C.

In reality, how can an academic library institute performance measures? This article describes why and how the Gelman Library initiated a program of performance measures, how we measured the accessibility of collections and services, and my assessment of the experience.

Performance Measures: The Practice

Why did the library institute performance measures?

Use and user studies have been conducted in Gelman Library for a variety of purposes for years, but the decision to create an ongoing program of performance measures emerged as a result of our formal planning process in 1986. Believing that the library needs to be a dynamic, change-oriented service organization, the university librarian introduced a strategic planning process. One critical element of this process is the environmental scan, which requires that we understand both our external and internal working environment. In part, a management information system helps to describe our internal library environment. The administrators in this library conceive of performance measures as part of that management information system. As we change policies and reallocate resources in order to accomplish our strategic goals, data from performance measures will reflect the results— both intentional and unintentional — of many of our planned changes.

Realizing that we could not allocate the necessary staff to conduct performance measures for all activities at once, we categorized activities and assigned priorities. Then we scheduled the implementation of measures in each category to be accomplished over a five year period. Categories of activities were designated as follows: accessibility of services and collections; collection quality; human resources; facilities; user education; library as gateway; and planning process. Although we roughly grouped library activities in these categories at the time we established the timeline, refinements are made as we address each one.

For example, we defined accessibility of services and collections to include in-house collection availability, turnaround time on interlibrary loan transactions and on searches for unfound items, utilization of equipment, and length of lines at service points.

Except for the accessibility category, our timeline relates to major events in the predictable future. For example, collection quality assessment began during the year when the Library played its first significant role in academic program review. Accessibility studies were selected as our first category because we had specific questions we wanted to answer, and because we wanted to learn more about the pattern of use by our primary user group as compared to that of visitors who make up a significant proportion of our user community. To understand our concern and why we believe that performance measures provide much better management information than does our professional judgment by itself, some information about this library will be helpful.

Gelman Library is the main university library on the main campus of The George Washington University. Our primary user groups, and therefore our target audience for collections and services, are the faculty, students, and staff of this University and, to a slightly lesser degree, the students and faculty of other member universities of the Washington Research Library Consortium. However, the campus is located in downtown Washington, D.C., adjacent to the Federal office area and easily accessible to more than seven hundred consulting firms and law firms. Unlike many urban university libraries, Gelman is available for on-site use by any member of the public who presents current photo identification at our registration desk. As a result, researchers from government agencies and private firms form a significant non-target clientele. Because the majority of our students are graduate students and most of them are employed in local government agencies or private firms, the "visitors" are not readily distinguishable from the students and faculty. This inability to differentiate at a glance complicates our ability to make informal assessments of how well we serve our primary clientele. Just to make things really challenging, we share a building with a number of administrative and academic offices and classrooms. Unfortunately, all of these non-library activities are accessible only through the Library's main entrance.

Selection and implementation of performance measures

The first and most essential step in establishing a performance measures program is educating the staff. Unless the staff understands and buys into the process, measurement of library activity is likely to be viewed negatively. We are so accustomed to thinking in terms of goals and standards and so accustomed to one-shot surveys, that it is difficult to accept the concept that measurement done consistently over time and done independently of standards will be valuable. The educational effort in Gelman Library had several components. One was an addendum to the strategic planning document which described the measurement and assessment model we would implement. That model defines measurement — as distinct from assessment — and lists the components of the process that pertain to each. Every staff member received a copy of the plan, including the addendum, during a staff gathering in the fall when the university librarian explained the reasoning behind the various provisions of the plan. In addition, articles about measurement appeared in our weekly staff newsletter. The most concentrated educational activity was the management retreat, which was attended by all administrators and heads of library departments and units. This one-day retreat focused on performance measures, with a short session on statistical reports that we file with local, regional, and national bodies. Because these reports tend to include primarily input data (budget, number of staff, and other resources) and very little output data (performance data such as reference statistics, loan transactions, etc.), dealing with the two topics in one retreat helped to clarify how performance measures differ from the data libraries traditionally collect. Aided by a specialist in educational measurement, we used the retreat as a workshop to learn the concepts and some techniques of measurement. As a result of the retreat, key staff members were able to imagine the usefulness of measurement in their own decision-making.

The first and most essential step in establishing a performance measures program is educating the staff.

During the 1987/88 academic year, I identified the kinds of studies that would tell us how successfully users actually locate books in our library, the length of lines at service desks, and whether or not we have sufficient equipment to provide access to the collections. Because we defined accessibility in its broadest terms, the equipment usage we studied included our catalog (which was on compact disc), indexing and abstracting services on compact disc, microform readers and printers, photocopy machines, and elevators. In the beginning, I drew on the published literature, experience, and a somewhat similar study conducted by Tracy Casorso in Gelman Library two years previously. Then I worked with a number of individuals and groups to design, plan, and implement the studies.

I sought two sources of expertise. One was statistical; the other was operational. A professor of management science and psychology provided the statistical and research design assistance. He offered invaluable advice about sampling, validity and other technical concerns. Most of all, however, he gave down-to-earth practical advice. He reassured me that studies done for purposes of management decision-making are quite different from experimental or laboratory research, where conditions can be controlled. Because our research is done in the real world with real library users (who may or may not be cooperative), where all kinds of events beyond our control influence human behavior, we need to note the events that may affect the results of our study. But those events do not invalidate the study. For example, if an exam in a large music class is scheduled for the day after our randomly chosen survey day, the use of audio equipment in the Media Resources Unit will be abnormally high. That will not be a "typical" day in that unit, but it isn't atypical either, so we note the cause of the high volume of use and include the data in the study.

The other source of expertise was the Gelman Library staff, the people who intuitively judge demand for services and adjust staffing levels accordingly. Not only did they provide a list of questions they hoped our performance measures would address, but they also gave thoughtful consideration to the selection of sampling time periods, design of data collection forms, and logistics. Because staff in this library work together in groups continually, it was easy to fit planning of performance measures into regular meetings of librarians, mid-level managers and supervisors, heads of service units, and so forth.

We planned data collection with the conviction that there is no such thing as a "typical week" in our library. There are, however, typical patterns within a week. For example, the usage patterns seem to be very similar on Monday through Thursday evenings. We identified nine such periods. Then we randomly selected nineteen dates during the fall 1988 semester for data collection, ensuring that we had sufficient representation of every survey period so that our survey samples would yield meaningful data. During the following spring semester we started a little earlier and were able to survey on twenty-two days.

In preparation for the surveys, we hired staff who would conduct the observations. We also developed and tested data collection forms for each study. One form, to be given to people using the serials lists, asked the users to note which journals they were seeking and whether or not they found the journal. Another asked users of the library's catalog to give the same information about the books they sought. Another set of forms was used by observers who walked through the library noting which machines were in use, which were out of order, how many staff members were working at specific desks, how many people were being assisted by those staff members, and how many people were waiting. Turnaround time on interlibrary loan requests, book search requests, and waiting time for appointment services could be derived from information noted on the normal request forms. Separate forms were designed for data collection at service desks, although these tended to be expansions on the data forms the staff routinely use.

Because we wanted to distinguish current GW faculty, staff, and students from alumni (a significant user group), from consortium faculty and students, and from all other researchers, we purchased labels in four colors to issue to library users as they entered the building. The color of the label indicated the individual's user category — GW user, consortium member, alumnus, or unaffiliated researcher. As individuals requested assistance at service desks or were observed using collections, library staff who collected were able to record transactions by category of user without having to ask each person about his/her affiliation.

On survey days the entrance staff, with assistance from additional staff during peak periods, handed each entrant a colored label and asked him/her to wear the label in order to help us conduct library surveys. Meeting some resistance by users who did not want to wear the label, on the second day we began offering a letter explaining the purpose of our surveys and the importance of wearing the labels. Over time we found that a large sign explaining the meaning of the various colors of labels answered most users' questions. As the survey progressed, people who were going to non-library portions of the building or just to study rooms declined the labels. But others wore the dots or presented them upon request as they sought assistance at service desks or when the observers made their rounds to record use of equipment and length of lines.

Most of our studies did not require conscious participation by library users. Library staff collected data through observation or as a routine activity during normal transactions at service desks. A user was conscious of being studied only if he/she failed to wear the colored label and, therefore, was asked to show the label to the data collector.

The only data collection that required conscious user participation was the collection use study, in which we asked people to note the books and journals they sought and whether or not they found the items. Users' willingness to fill out (or submit) the worksheets varied from modest to poor. As a result, while we received sufficient response to draw general conclusions about the causes of user failure to find the materials they sought, the decline in response rate over the course of the semester prevented us from answering some of our more specific questions. For example, we had wanted to know whether the causes of user failure varied by time of semester. The number of survey responses dropped as the semester progressed, leaving us with insufficient data to analyze variation by time of semester.

Usefulness of the measures

In this initial set of studies, we collected a great deal of baseline data that was useful in documenting demand for specific services by category of clientele. Many of our assumptions about usage patterns were confirmed, and some of our assumptions about our shortcomings were disproved. For example, we had believed that we had long lines waiting at photocopy machines and that unaffiliated users were tying up our *ABI Inform* stations. Neither of these turned out to be true. As a result, we decided not to purchase additional copiers and postponed implementation of measures to restrict use of the selected reference tools on compact disc. The impact of our malfunctioning circulation computer system and the crowded conditions of our stacks could be described objectively and quantitatively as a result of the collections use study. We could state with confidence that we have sufficient access tools of various types to meet users' needs, except at peak demand periods, and could identify the times and places where we most feel the impact of unaffiliated users. As a result, we have changed some service hours, changed some policies and practices (such as providing priority service to GW members who present identification at the Reference Desk), and identified improvements we would like to make if the opportunities arise. But most importantly, staff members who participated in data collection have a new awareness of the usefulness of performance data for decision making. Finding that the data disproved some of our assumptions provided a good demonstration of the need to base decisions on hard data mixed with experience and intuition.

In the 1990/91 academic year we will repeat some of the accessibility studies to determine whether the deselection process (which loosened up space in some stacks areas), a new circulation system, staffing reallocations, and some policy changes have had the desired effects. When we do that, the full usefulness of performance measures to record changes over time will be demonstrated. Meanwhile, we have proceeded with planning and implementing performance measures for other library activities.