

Title: *Making Incremental Improvements to Public Library Comparative Statistical Practices*

Ray Lyons and Jason Holmes

ABSTRACT

The majority of U.S. public libraries have yet to immerse themselves in the “culture of assessment” that library researchers have promoted in recent years. For the most part, these libraries dutifully collect and report traditional library input/output measures to state and funding authorities. Some libraries might devote time to comparing their statistics to “peer” libraries. They might use composite measures, such as national public library ratings, or review comparative data from state libraries or from the Public Library Association, or utilize tools developed by the U.S. National Center for Educational Statistics.

These types of statistical comparisons—composite national ranks and selected local comparisons—suffer from methodological weaknesses that have not been completely explored by the library profession. This paper examines the shortcoming of comparative approaches: the non-equivalence of activities that traditional “counts” represent, the imperfect selection of peer libraries, the lack of criteria for judging adequate performance levels, the psychological attraction of the *More-Is-Better-Myth*, and certain misconceptions concerning justifiable interpretations of library statistics. The paper suggests a few modest improvements to mitigate the deficiencies of these comparative methods. In addition, it advocates for re-thinking the basic premises of evaluating library statistical and comparative data.

Making Incremental Improvements to
Public Library Comparative Statistical Practices

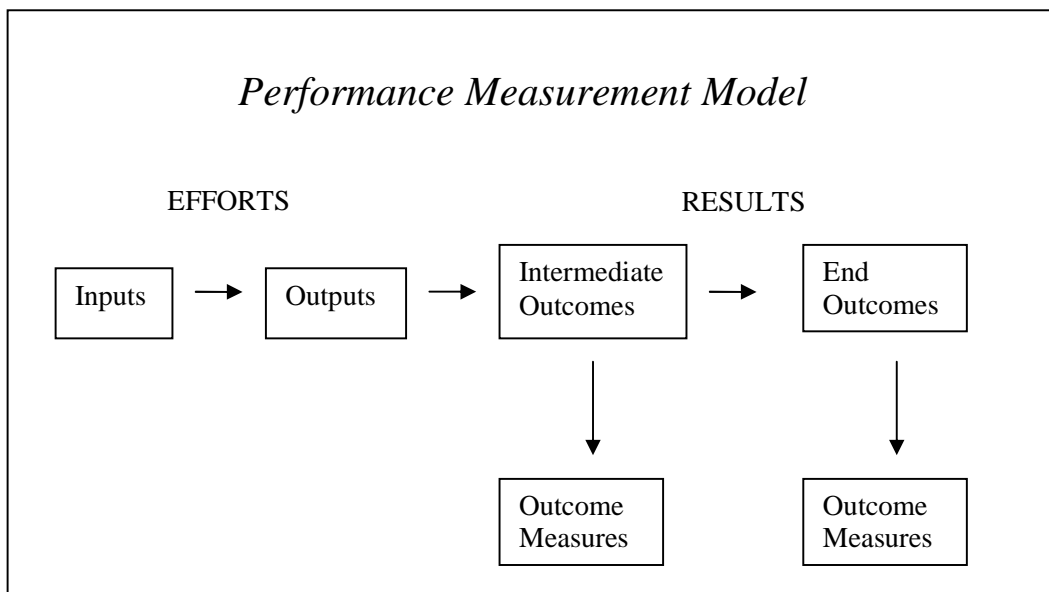
Ray Lyons and Jason Holmes

Paper presented at the
Library Assessment Conference:
Building Effective, Sustainable, Practical Assessment

Seattle, Washington
August 4–7, 2008

The purpose of this paper is to consider modest solutions to problems with making comparisons of public library performance using statistical data. Before describing these problems and solutions, it will be useful, first, to consider the context within which the practice of comparative library statistics operates, and, second, to describe how public libraries currently perceive these practices.

Organizations use statistical indicators as part of a rational planning and management process referred to alternatively as *performance measurement*, *performance assessment*, and *performance monitoring*. A basic performance measurement model is depicted in Figure 1. Ideally, a main objective of the approach is to use information about organizational performance to inform management decision-making and improve organizational performance and effectiveness. For public and not-for-profit organizations, performance measurement is also intended to generate information for accountability purposes, that is, for reporting program results to funding sources and to the public-at-large.



Based on a general systems theory model, performance measurement envisions organizations as production systems where *inputs* are transformed into *outputs* that interact with the organization's environment to produce *outcomes* in both the short-term (*intermediate*) and long-term (*end*). *Outcome measures* are indicators collected in order to determine the degree to which an organization's intended outcomes (and, in some cases, unintended outcomes) have been achieved.

Over the past thirty years this performance measurement model has become the mainstay of collecting public library statistics in the United States. The prime impetus for the establishment of this approach has been the promotion of *results-oriented planning*

and management by the Public Library Association (PLA).¹ Statistical data collection for local assessment of library performance was most notably promoted in the PLA instruction manual by Van House et al.² Following PLA's lead, Van House and her colleagues downplayed pre-established library standards in favor of standardized data collection coupled with customized local data interpretation. They encouraged libraries to collect statistics for "assessing current levels of performance, diagnosing problem areas, comparing past, current, and desired levels of performance [and] monitoring progress toward the library's mission, goals, and objectives."³

To reiterate, local libraries alone would be responsible for determining standards against which performance statistics would be evaluated. The libraries were also obliged to utilize the statistical data as material for the process of organizational self-evaluation. In addition, the PLA espoused routine comparison of each library's statistical data with other similar ("peer") public libraries. Theoretically, comparisons would provide a library with further indications of the acceptability of its own performance.

Public Library Perceptions and Utilization of Statistical Measures

As an indication of how public libraries currently perceive and utilize comparative library statistics, a study of Ohio libraries recently conducted by the primary author (Lyons) is briefly reviewed here.⁴ This exploratory study surveyed 90 public libraries selected via a stratified random sample of all public libraries in Ohio. Forty-two libraries (47%) responded, most via the study's online questionnaire. Seven libraries participated in brief telephone interviews conducted by the primary author. Further details of the study methodology are available in the full study report.

One questionnaire item in the survey concerned how often library management teams review standard library statistics, regardless of whether they make comparisons with other libraries. Table 1 indicates the frequency with which responding libraries examine their own input statistics. Nearly two-thirds of reporting libraries review both operating and print materials expenditures monthly. Other categories like print material counts, subscriptions, and so forth are reviewed annually.

Output statistics that Ohio public libraries say they review periodically are shown in Table 2. Eighty three percent of the libraries report reviewing circulation data monthly, while two-thirds review interlibrary loan, internet terminal use, and library website usage measures on a monthly basis. Reference transactions are reviewed either quarterly or monthly by 58% of the libraries. In-house use of library materials is reviewed annually or more often by more than 60% of reporting libraries.

The survey also asked respondents whether, in the past two years, they had made comparisons of their own statistics with those of other libraries. About 6% of the libraries had not made these comparisons. Figure 1 presents the frequency with which the 32 libraries who do make comparisons at least biennially compare their statistics with those

of other libraries. Seventy-eight percent of these libraries say that they compare their statistical data semi-annually or annually, and 19% report doing so quarterly or monthly.

<i>Statistical Indicator</i>	<i>Annually</i>	<i>Quarterly</i>	<i>Monthly</i>	<i>Weekly</i>	<i>Rarely</i>	<i>Not Sure</i>
Operating expenditures	12%	9%	65%	15%	0%	0%
Print Mat. Expenditures	12%	18%	62%	9%	0%	0%
Electronic Mat. Expenditures	39%	21%	36%	3%	0%	0%
Print Materials	56%	9%	29%	3%	3%	0%
Print Subscriptions	74%	12%	15%	0%	0%	0%
Audio/Video Materials	58%	9%	30%	0%	3%	0%
Databases	63%	14%	6%	6%	6%	6%
Internet terminals	51%	11%	14%	9%	6%	9%
FTE Staff	79%	6%	12%	0%	3%	0%

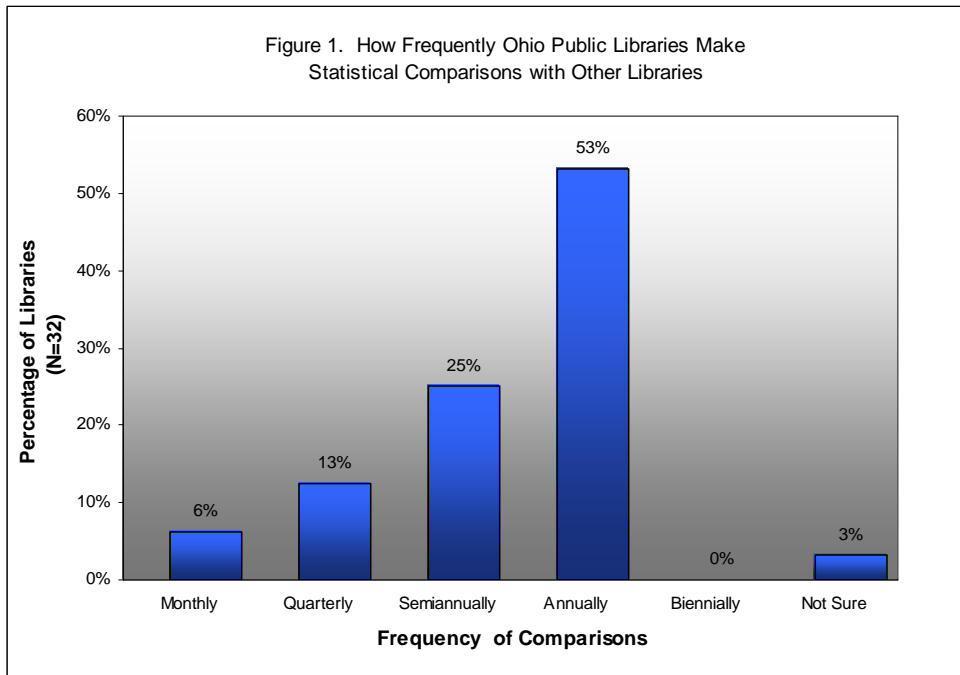
Source: Lyons (2008)

Another survey question polled libraries about their identification of peer libraries when making comparisons. Their responses appear in Figure 2. The table indicates that nearly 75% of the libraries made selections based on population or demographic variables. Seventy-seven percent of the libraries select libraries having statistical indicators that are similar to the library's own data. Six percent of the libraries say that they select peer libraries based on programming and service offerings.

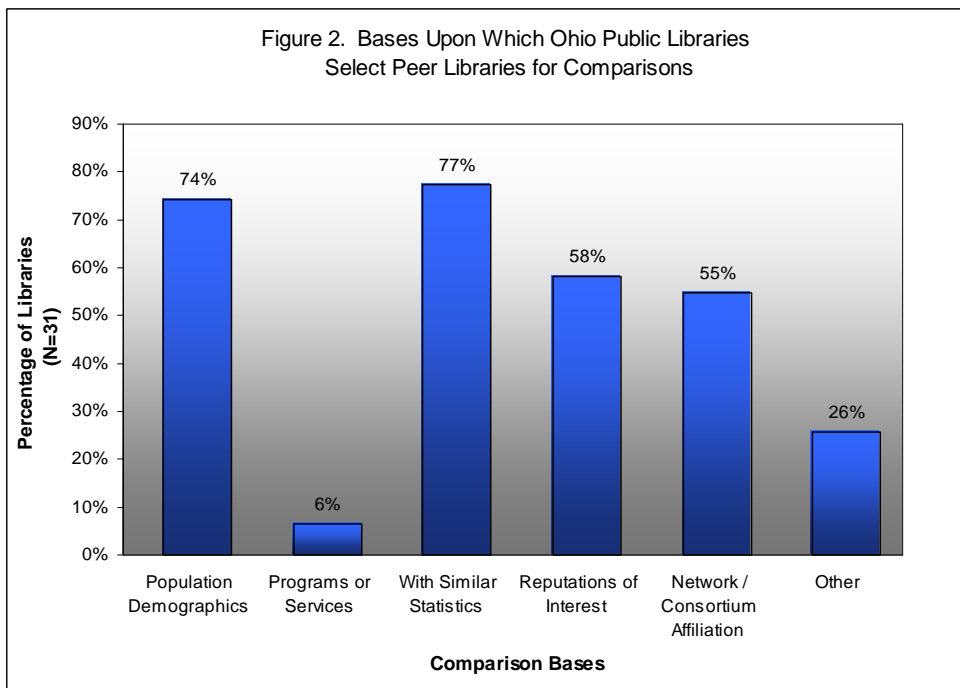
<i>Statistical Measure</i>	<i>Annually</i>	<i>Quarterly</i>	<i>Monthly</i>	<i>Weekly</i>	<i>Rarely</i>	<i>Not Sure</i>
Circulation	9%	0%	83%	9%	0%	0%
In-house Mat. Use	21%	12%	26%	3%	34%	0%
Interlibrary loan	23%	6%	66%	6%	0%	0%
Visits	21%	15%	54%	6%	3%	0%
Reference Transactions	32%	24%	34%	0%	9%	0%
Program attendance	32%	9%	51%	6%	0%	0%
Electronic Mat. Use	19%	6%	60%	0%	9%	0%
Internet Terminal Use	15%	6%	66%	3%	3%	3%
Website Use	12%	9%	63%	0%	9%	3%

Source: Lyons (2008)

For libraries reporting making comparisons at least biennially, Figure 3 indicates which statistical indicators the libraries review in these comparisons. Ninety percent or more of the libraries use three main statistical indicators to make comparisons with other libraries—material expenditures, total operating expenditures, and circulation. All of the libraries make comparisons using material expenditures. Ninety-seven percent say they

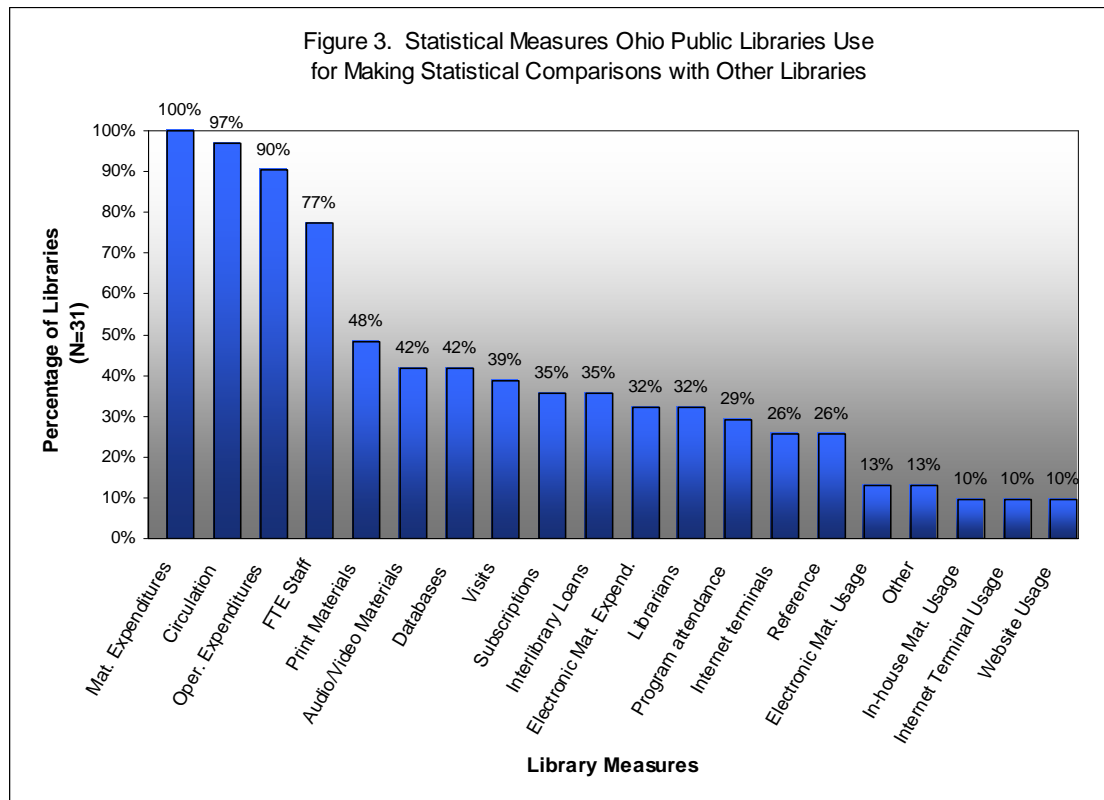


Source: Lyons (2008)



Source: Lyons (2008)

make comparisons using circulation, and 90% use total expenditures. Staffing statistics are used for comparisons by nearly 80% of the libraries. (Open-ended questionnaire comments and interview responses suggest that these comparisons may be for salary studies rather than for performance assessment.)



Source: Lyons (2008)

Key Problems with Library Statistics

The exploratory survey just described did not delve into how libraries interpret their operational statistics, either alone or in comparison with other libraries' statistical data. As already mentioned, PLA has recommended that local libraries determine the meaning of each library's statistical indicators in the context of the library's mission, objectives, community characteristics, chosen service responses, and so on. Van House, Weill, and McClure also note that:

There are no "right" or "wrong" scores on an output measure; "high" and "low" values are relative. The scores must be interpreted in terms of library goals, scores on other measures, and a broad range of other factors.⁵

We can presume that Van House et al. would have also applied this admonition to input measures had PLA not omitted these from its measurement model.

Fourteen years prior to the publication of PLA's *Output Measures* manual, library researchers had already expressed doubts about the utility of comparative library statistics. Reflecting on her participation in a classic public library performance research study,⁶ Altman writes:

The project team was philosophically opposed to the practice of standard comparisons [of libraries using input/output measures] because of the arbitrary way in which they were set and the general lack of care used in making the comparisons. Had we taken it upon ourselves to pronounce that certain numbers were "good" or "bad," we, too, would have been rightly accused of being arbitrary...The study team felt strongly...that each library staff should decide for themselves whether the findings for that library were acceptable in terms of performance expectations.⁷

Here Altman is calling attention to the first, and perhaps most important, of several problems with using library statistics: There are no established standards for evaluating the data. Specifically, there is no objective way to determine whether the magnitude of an input statistic represents high, medium, or low levels of resources, or whether the magnitude of a library output statistic represents excellent, satisfactory, or poor performance. Yet, as Rossi and others in the field of program evaluation advise, standards of some sort are necessary to draw evaluative conclusions about program and organizational performance.⁸

Nevertheless, thanks to the efforts of the U.S. National Commission on Libraries and Information Science, the National Center for Educational Statistics, state library authorities, the PLA, and others, libraries in the U.S. can take advantage of rather sophisticated systems for gathering standardized library statistics. The ready availability of these standardized statistics now makes the mechanics of library comparisons practical and convenient.⁹

Standard definitions used by these collection systems, however, introduce another key problem with these statistics. The act of standardizing library attributes and activities—creating what statisticians call *classes of equivalence*—limits the meaningfulness of library statistics. Establishing these standard classifications requires that numerous objects, events, and other library attributes be viewed as equivalent. All books, periodicals, staff, visits, circulation transactions, reference questions, programs, and so forth are assumed to be exactly alike for purposes of statistical counting. This assumption ignores essential differences among these entities, including differences in complexity, sophistication, relevance, quality, worth, effectiveness, efficiency, and significance.

Put another way, standardization *homogenizes* the information content of the statistics, *reducing* the underlying phenomena to an arbitrarily small collection of facts. The resulting data provide only the most broad-brush indications of library performance. General uses for data with such abridged content may well be appropriate for regional, national, and international planning and evaluation. On the other hand, the data are not

informative enough to provide complete or accurate descriptions of individual library performance.

Prior to the appearance of total quality management and benchmarking in the private sector in the U.S., PLA had encouraged library comparisons as an important use for standardized statistics. Regrettably, PLA was as vague about the methods by which libraries could identify peer libraries as they were about techniques for drawing conclusions from the statistics. In the meantime, tenets of the quality movement drew the attention of local and state government administrators. Eventually, *comparative performance measurement* appeared in the form of *municipal benchmarking* as a tool to “place local performance in context and, where major performance gaps are detected...suggest the need for additional analysis.”¹⁰ Even though municipal benchmarking relies heavily on cross-jurisdictional performance standards, as a management practice it serves to reinforce PLA’s emphasis on library comparisons. Presently, benchmarking is considered a respectable management practice by libraries in the U.S and internationally.¹¹

Despite its popularity, library benchmarking efforts are hardly straightforward and sometimes frustratingly inconclusive. International library statisticians Poll and te Boekhorst offer numerous caveats to libraries planning to make statistical comparisons.¹² They advise libraries to carefully examine all possible explanations for measurement variances. Other proponents of comparative performance measurement acknowledge that the practice can produce misleading results. Ammons notes that localities having high performance statistics can still be neglecting particular constituent populations, and that local statistics are self-reported, unaudited, and often inaccurate.¹³

Beyond these interpretational difficulties, benchmarking has a major shortcoming reluctantly acknowledged by its proponents. This is the lack of effective methods for identifying suitable peer organizations. Morely, Bryant, and Hatry conceded that:

no two...jurisdictions or organizations are completely comparable. Each has unique characteristics. As a result, it is impossible to find organizations that are exactly comparable.¹⁴

The romantic image of chemists centuries ago weighing substances on a balance comes to mind. Only after confirming the parity of chemical samples could the scientists proceed to conduct their experiments. Today we have no analogous measurement tools to establish parity between organizations being compared. Yet, without equivalent peer organizations, benchmark comparisons will be gross estimates at best. At worst, the comparisons can be inaccurate and invalid.

Another problem with standardized library statistics is imprecision in collected data. This imprecision is due to inconsistent collection methods, clerical mistakes, sampling error, “gaming” (the intentional production of data to produce desirable results), statistical imputation (statistical estimation of missing data), and other factors. In individual library comparisons, libraries may be able to investigate irregularities in the

data, if they are fortunate enough to recognize them. However, aggregate comparisons such as state or national rankings on specific library indicators are often published without adjustments or corrections for imprecision in the data, nor with clear disclosures concerning these sources of inaccuracy.

Related to data imprecision is the lack of audit mechanisms for verifying library self-reported operational statistics. Library accounting records are subject to professional audits to confirm that financial statements represent the true financial standing of the library. However, there are no analogous audits of library operational data. So, it is impossible to determine whether or not the statistics accurately represent a library's performance.

Perhaps the most intriguing aspect of library statistics is the psychological tendency to view higher numbers as favorable and lower numbers as unfavorable. As already noted, performance statistics are not automatically indicative of performance successes or failures. Ideally, they need to be evaluated according to well-thought out standards. As also noted, the library profession lacks these standards. Consequently, libraries tend to rely on *magnitude* itself as an indication of success. One might call this the *More-Is-Better Myth*. Even so, as Van House et al. note, "More activity does not necessarily mean better activity."¹⁵ More importantly, mere quantities tell nothing about the extent to which community needs have been met (an issue we address in more depth below). And it is quite feasible that efforts to increase the quantity of library services can result in a lowering of service quality.

While standardized statistics are fairly simple to conceptualize and collect, their interpretation is surprisingly complex. Van House, Weill, and McClure suggested that library measures:

...reflect the interaction of users and library resources, *constrained by the environment in which they operate*. The meaning of a specific score on any measure depends on a broad range of factors including the library's goals, the current circumstances of the library and its environment, the users, the manner in which the measure was constructed, and how the data were collected.¹⁶
(italics added)

In many cases a library's performance reflects its environment as much as its operational effectiveness. Some public libraries enjoy especially fertile economic, demographic, or political settings that serve as catalysts to library performance, while others do not. Yet, this important aspect of organizational effectiveness is not captured in library statistics.

By recounting this litany of problems and challenges with library statistics, we do not mean to imply that the PLA and other leaders in library statistical policy have failed to appreciate the complexities of library performance measurement. Nevertheless, it is fair to conclude that current library statistical policies have been ineffective, for the most part. It was unrealistic to expect that public libraries would be able to interpret and utilize statistics productively. Other than making comparisons with their own historical data,

most libraries publish operational statistics without any interpretation of their adequacy or appropriateness. Neither do libraries seem to have a good understanding of conclusions that can justifiably be drawn from these data.

General misconceptions about formulating meaningful interpretations from library statistics are epitomized by recent advocacy campaigns of the American Library Association (ALA) and other library organizations. The ALA, for example, has endorsed a curious array of statistical comparisons as collections of “quotable facts.” Consider facts such as these:

*There are more public libraries in the U.S. than there are McDonalds.*¹⁷

*In the state of West Virginia public libraries are more numerous than Wal-Marts by a factor of more than 5 to 1.*¹⁸

*If the number of reference question asked weekly in U.S. public and academic libraries were represented by an end-to-end line of questioners, the line would stretch from New York City to Juneau, Alaska.*¹⁹

These comparisons become nonsensical when incompatible units of measure are mixed, as when circulation counts (items) are compared to fans (people) watching the 2006 Super Bowl game, and reference counts (queries) exceed the population (people) of Galveston, Texas.²⁰

It is obvious from statistical comparisons like these that our profession is at a loss for meaningful ways in which to evaluate library statistics. Probably, the professionals who compiled these brochures do not realize where this type of thinking ultimately leads. These portrayals idolize quantification over meaning, adopting by default the *More-Is-Better* stance mentioned earlier. This strategy, though, quickly evolves into grasping for sensational straws in order to give meaning to measures that might otherwise be indecipherable to the general public.

The basic principles of performance measurement can assist us in critiquing these kinds of “facts.” These principles lead us to ask questions like: Does the American public want or need more libraries than it wants or needs McDonalds? Than it wants or needs Wal-Marts? Does providing reference answers to a chain of humans stretching from New York to Juneau tell us how excellently, satisfactorily, or poorly libraries have performed? What if libraries actually spend too much time on reference questions, time that could be better spent addressing a wider array of information needs in our communities? If library advocates considered questions like these, their campaigns would become more credible and relevant.

Incremental Improvements for Library Comparative Statistics

At least informally, public libraries already recognize many of the inadequacies of standardized library statistics that we have outlined. However, it would be beneficial for libraries to acknowledge these challenges more publicly, for instance, in reports to boards of trustees, funding authorities, and to the public at large. Continual reminders that these data are limited and incomplete reflections of library performance might prompt the development of better measurement tools. When libraries do publish standard library statistics, or make comparisons using these, they should also be obliged to include other types of performance data to complement the statistics. This too will serve as a reminder that standardized statistics alone are insufficient.

When public libraries intend to make statistical comparisons with other public libraries, these peer libraries need to be selected based on a variety of relevant library attributes. At a minimum, only peer libraries that can be shown to be equivalent in population size, community demographics, service response choices, and budget should be included in any comparisons. If a library is unable to identify peers based on these four attributes, *then no comparisons should be made*. Alternatively, libraries may select one or two of their own services responses—or services or programs—and search for potential peer libraries based on these only. For example, consider two suburban libraries whose demographics are alike, but whose population counts differ significantly, and who also both offer career services to their communities. The libraries could make comparisons of indicators relevant to these programs only. In this case, it would also be important for libraries to collect demographic data about program participants in order to assure that the same general clientele were being serviced in both libraries.

Another area ripe for improvement is libraries' understanding of relationships between statistical indicators, and of managerial decisions that can affect one statistical indicator or another. Many libraries understand the relationship between circulation policies and circulation counts, since circulation has traditionally been such a highly visible indicator. Along these same lines, libraries should explore relationships between other measures, for instance, of changes in registered patrons and circulation, or increases or decreases in website usage and circulation or renewal rates.

Using a one-size-fits-all list of statistical indicators is counterproductive to performance measurement in public libraries. Smaller public libraries need a different set of core measures than larger libraries, as do rural libraries as opposed to urban or suburban libraries. The same is true for inner city libraries where, for instance, circulation counts are hardly as important as visits or programming. Similarly, with additional requirements for libraries to report electronic resource usage, difference between libraries need to be noted. Again, inner city and rural libraries will not be able to report usage comparable to suburban libraries.

Even though interpretation is essential to making productive use of library statistics, this task is often glossed over in discussions of performance assessment. Here interpretation refers to conclusions about library performance drawn from standardized

statistics. Library data reflect quite concrete events and resources, such as patron behavior (visits, circulation), materials available, money expended, and so forth. Consider a library statistic utilized internationally—how many *seats* a library has. Literally, this statistic represents the number of chairs in a library. We might assign a slightly more abstract meaning to the statistic by saying it also indicates the extent to which a library encourages in-house use of materials, browsing, and so forth. We could also interpret the number of seats as a reflection of a library's interest in accommodating persons with disabilities. Or we might say the statistic reveals the extent to which the library might be perceived as a comfortable and welcoming place.

This example illustrates that statistical data can (and will) be translated into more abstract concepts, beyond the literal meaning of the data. It is in this area that libraries must be particularly careful to avoid drawing unreasonable conclusions. For example, some library advocates promote using visit counts as reflections of library *value*, even though the statistic indicates merely how many people enter library buildings. A more reasonable interpretation is the ability of the library to attract in-person attendance. At the same time, it is incumbent upon libraries to provide some type of interpretation of their statistical data while avoiding making exaggerated claims that the data cannot support.

Beyond these modest suggestions, there is one rather monumental improvement that the library profession will ultimately need to contemplate. The current paradigm of comparative performance measurement needs to be changed. Rather than evaluating performance based on comparisons with peer libraries (or even a library's historical statistics), libraries should be judging their types and levels of service delivery based on verified levels of community needs. Thus, the most fruitful performance measurement efforts that public libraries can undertake are *determining ways to assess levels and constellations of community needs and assuring that amounts and types of library services correspond closely with these needs*. This is the only way to satisfactorily answer the question of whether a given level of service is excellent, good, or poor.

As an inspiration to take up this challenging paradigm shift, we close with this quotation by leading program evaluation theorist, Michael Scriven:

[Performance] monitoring and ... goal achievement evaluation tell the program manager what he or she needs to know about the progress of the program toward its goals, but they do a severely limited job for the recipient, the taxpayer, the citizen, and those concerned with the welfare of program impactees [sic]. Most consumer or taxpayer groups have little interest in whether a program meets its goals as such, *only in whether it does something that needs doing, whether the cost is reasonable, and whether it does it better than alternative ways of doing it*. And, from their point of view, the standards for judging whether it does the job well or better must include standards of equity...not just short-term economy. None of these issues is addressed by goal achievement evaluation, and only their surface is scratched by compliance monitoring.²¹ (italics added)

NOTES

¹ See for instance: Sandra S. Nelson, *The New Planning for Results: A Streamlined Approach* (Chicago: American Library Association, 2001).

² Nancy A. Van House et al., *Output Measures for Public Libraries: A Manual of Standardized Procedures*, 2nd ed. (Chicago: American Library Association, 1987).

³ *Ibid.*, 1.

⁴ Ray Lyons, *Use and Perceptions of Comparative Statistics by Ohio Public Libraries* (2008) <http://www.plstatreports.com/Compare/UsePerceptComparStats.pdf>.

⁵ Nancy A. Van House, Beth T. Weill, and Charles R. McClure, *Library Performance: A Practical Approach* (Chicago: American Library Association, 1990), 7.

⁶ Ernest R. De Prosopo, Ellen Altman, and Kenneth E. Beasley, *Performance Measures for Public Libraries* (Chicago: Public Library Association: 1973).

⁷ Ellen Altman, "Reflections on Performance Measures Fifteen Years Later," in *Library Performance, Accountability, and Responsiveness: Essays in Honor of Ernest R. De Prosopo*, ed. Charles C. Curran and F. William Summers (Norwood, NJ: Ablex: 1990), 13.

⁸ Peter H. Rossi, Mark W. Lipsey, and Howard E. Freeman, *Evaluation: A Systematic Approach*, 7th ed. (Thousand Oaks, CA: Sage: 2004), 171-172.

⁹ See for instance: *Compare Public Libraries*, <http://harvester.census.gov/imls/compare/index.asp>.

¹⁰ David N. Ammons. *Municipal Benchmarks: Assessing Local Performance and Establishing Community Standards* (Thousand Oaks, California: Sage, 2001), 7.

¹¹ See for instance: Leo Favret, "Benchmarking, Annual Library Plans, and Best Value: The Implications for Public Libraries," *Library Management*, 21, no. 7 (2000): 340-348; Roswitha Poll, "Benchmarking with Quality Indicators: National Projects," *Performance Measurement and Metrics* 8, no.1 (2007): 41-53.

¹² Roswitha Poll and Peter te Boekhorst, *Measuring Quality: Performance Measurement in Libraries* (Munich: K.G. Saur, 2007).

¹³ David N. Ammons. *Municipal Benchmarks*.

¹⁴ Elaine Morley, Scott P. Bryant, and Harry P. Hatry, *Comparative Performance Measurement* (Washington, DC: Urban Institute, 2001), 6.

¹⁵ Van House et al., *Library Performance*, 7.

-
- ¹⁶ Ibid.
- ¹⁷ American Library Association, *Quotable Facts about America's Libraries*, <http://www.ala.org/ala/issues/toolsandpub/quotablefacts/quotablefacts.cfm>.
- ¹⁸ West Virginia Library Association, *Quotable Facts about West Virginia Libraries* http://www.ala.org/ala/issues/toolsandpub/quotablefacts/WestVa_Web.pdf.
- ¹⁹ American Library Association, *Quotable Facts*.
- ²⁰ Texas Library Association, *Quotable Facts about Texas Libraries*, http://www.ala.org/ala/issues/toolsandpub/quotablefacts/Texas_web.pdf.
- ²¹ Michael Scriven, *Hard-Won Lessons in Program Evaluation* (San Francisco: Jossey-Bass, 1993), 19-20.