LIBRARY IMPROVEMENT THROUGH DATA ANALYTICS

LESLEY S. J. FARMER
ALAN M. SAFER

Neal-Schuman
An imprint of the American Library Association
CHICAGO 2016

www.alastore.ala.org

© 2016 by the American Library Association

Extensive effort has gone into ensuring the reliability of the information in this book; however, the publisher makes no warranty, express or implied, with respect to the material contained herein.

ISBN: 978-0-8389-1425-0 (paper)

Library of Congress Cataloging-in-Publication Data

Names: Farmer, Lesley S. J., author. | Safer, Alan M., author.
Title: Library improvement through data analytics / Lesley S.J. Farmer, Alan M. Safer.
Description: Chicago : ALA Neal-Schuman, an imprint of the American Library Association, 2016. | Includes bibliographical references and index.
Identifiers: LCCN 2015046382 | ISBN 9780838914250 (print : alk. paper)
Subjects: LCSH: Libraries—Evaluation—Statistical methods. | Library administration—Decision making—Statistical methods. | Six sigma (Quality control standard) | Library statistics—Case studies.
Classification: LCC Z678.85 .F37 2016 | DDC 025.007/23—dc23 LC record available at https://lccn.loc.gov/2015046382

Cover design by Alejandra Diaz. Imagery © Shutterstock, Inc.

Text composition by Dianne M. Rooney in the Chaparral, Gotham, and Bell Gothic typefaces.

♾ This paper meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).

Printed in the United States of America

20 19 18 17 16 5 4 3 2 1

www.alastore.ala.org

Contents

www.alastore.ala.org

www.alastore.ala.org

PART I

Overview

www.alastore.ala.org

1

Introduction

I can't find the articles I need!
What does it take to get a new book on the shelf before it becomes old?
No one uses our self-checkout system.
Should we subscribe to ebooks?
Why are students just using Google and Wikipedia?
How do we show the value of the library to our organization?

These questions are just a sampling of concerns that library communities ask. Mindful of budget constraints, libraries have to optimize their operations in order to provide the most cost-effective services. At the same time, librarians need to respond to community needs to optimize clientele use and satisfaction. Increasingly, libraries are having to demonstrate their value, and data-driven decision-making is expected of library management. These sometimes conflicting demands may be daunting. Data analytics provides a tested way to assure continuous library improvement. This chapter explains the basis for data analytics, highlighting quantitative aspects, and states how

www.alastore.ala.org

it can benefit library management. A library case study concretizes the use of data analytics to show library value (Soria, Fransen, & Nackerud, 2014).

LIBRARY IMPROVEMENT

In the midst of putting out library brush fires, the idea of library improvement seems like a distant ideal. Just maintaining the status quo can be a library manager's mindset, especially during economic downturns. Perhaps the issue of library improvement comes to the fore during trustee meetings or mandated five-year plans. On the other hand, most librarians do seek to provide the best cost-effective programs of resources and services to meet the interests and needs of their communities. And there's always room for improvement. In fact, hard times require librarians to take a close look at their operations to make them more cost-effective and value-added in the eyes of their communities.

Continuous library improvement calls for an ongoing cycle of planning, action, monitoring, evaluation, and reflection. Although annual reports and major campaigns can serve as touchstones, attention to ways to improve libraries should occur routinely. At first glance, improvement might be envisioned as a bigger collection, more events with greater participation, increased staffing, and increased hours of operation. However, bigger is not always better. Moreover, library improvement can occur without adding resources, and even if resources are reduced. The ultimate criterion is increased value in terms of the best return on investment.

Management circles tend to focus on a subset of continuous improvement: continuous quality improvement (CQI). This ongoing process evaluates how organizations work, and has the intent of reducing waste, increasing efficiency, and increasing satisfaction. The underlying philosophy asserts that problems emerge from ineffective leadership, work design, or communication rather than from worker intent. The goal of CQI is stable and predictable process results, with reduced process variations. It should be noted, however, that achieving sustained quality improvement requires long-term commitment and effort from the entire organization.

DATA-DRIVEN DECISION-MAKING

In a climate of heightened attention to accountability, institutions and organizations are undergoing increasing scrutiny. Recent economic realities have also driven decision-makers to closely examine their operations' cost-effectiveness. Even without external drivers, libraries want to make good decisions. To that end, decision-makers need to gather and analyze meaningful

www.alastore.ala.org

data, and act upon it. What is the return on investment? The phrase "show me the data" rings loudly.

Such emphasis on data means that processes should have characteristics that are measurable so they can be analyzed, improved, and controlled. Some library functions lend themselves easily to such metrics, for example, information services and circulation. It is usually a straightforward process to count frequencies, reference-fill rate, and time-per-unit. However, softer services such as reader's advisory and instruction are harder to measure beyond the surface level (i.e., whether a person checked out a recommended book, or whether a student located an article). Even story hour attendance tells just part of the data story.

Although a list of assets (both human and material) and activities provides useful information about input, it does not suffice. Administrators and other stakeholders need to know how those assets and activities impact their clientele and community at large. Not only should data be collected about the use of those assets and generated products output, but data needs to be collected about the consequences of such use, such as increased literacy and income (outcome). For instance, libraries traditionally cite collection size and circulation figures, but they also need to ascertain if collection use translates into gains in reading ability, more job marketability, or other measurable outcomes. The result is system-based data-driven management.

One of the issues of this system-based data-driven approach is its open-endedness. Unlike a closed system where all parts are internally controlled, libraries operate in an open system where clientele come and go. Librarians cannot control those clients, either in terms of using the library or in leveraging their library experiences to improve themselves or contribute to society. Even something as elemental as checking out a book does not guarantee that the person will read the book, let alone apply the content read to his life. On the other hand, most people act because they think they will benefit from their action, even if it is to enter the library to get out of the rain.

Fortunately, librarians can take this independent action into consideration when making data-driven decisions. Librarians can do an environmental scan of their community to ascertain its status and needs. Census figures can give information about neighborhood demographics, health, income, educational attainment, industry sectors, and so on. School data also gives socioeconomic and demographic data. Local entities such as the Chamber of Commerce and realty companies also keep useful data. These data signal strengths that the library can leverage; for instance, if athletics brings money and attention to the community, then the collection should reflect that emphasis, and library programs can incorporate sports topics and guest speakers. If unemployment is high, then career services might be ramped up. The library can then collect baseline data about the library and community, and then improve its collection and service, re-measuring those data points to see if the library has

www.alastore.ala.org

contributed at least indirectly to the community's welfare. Thus, both internal and external data provide direction for library efforts, and can show the library's value to its clientele and community.

A CASE FOR QUANTITATIVE APPROACHES

Data take many forms: numbers, text, pictures, stories, products. Even politicians know that facts and figures can impress the head, and that a compelling anecdote can move the heart. At the end of the day, though, librarians need to make an evidence-based case for their library programs, and quantitative data need to form the basis for their stances.

Even at the descriptive level (such as identifying mean, mode, range, variance), data can inform decision makers. For instance, data might reveal that the public library is frequented most heavily in the afternoon, with half of the users being high schoolers. The library might well be poised to add reference staff and add YA novels to the collection. But first, librarians should find out the reason for such attendance by surveying users and interviewing local school librarians and other youth-serving agencies. Perhaps the school library closes immediately after the last class—can the public librarian advocate for longer school library hours, or cosponsor library programs? Perhaps few after-school activities are available in the community; could the public library partner with community agencies to do outreach activities or start interest and service clubs? Survey data might reveal that student library users tend to use the library as a place to hang out, which might lead to after-school library clubs; on the other hand, if the main purpose is to study, then the library might establish tutoring services. In any case, quantitative data provide the basis for further investigation and thoughtful analysis.

Descriptive statistics basically summarizes and describes data. Although descriptive data comprise much of library data analysis, the real power of data emerges when using more sophisticated inferential statistics to test hypotheses about populations and predict population characteristics. Inferential statistics can also explore relationships between factors at a more nuanced level. For instance, what is the relationship between library use and corporate profit? A positive significant correlation between these two variables can lead to assumptions about the library's value (although a causal relationship is much harder to prove). Factor analysis can identify the most important library services relative to student achievement. Regression analysis might inform librarians about optimum length of times for training.

Quantitative data not only provides more in-depth internal analysis, but it also enables libraries to benchmark with comparative institutions locally and nationally. With the same kind of data measured in the same way, libraries

www.alastore.ala.org

can find out if their budgets or staffing are equitable, for instance. Furthermore, factors such as socioeconomics can be held constant, so that libraries in disadvantaged neighborhoods can still show that they make a difference in their communities relative to their high-income suburban counterparts.

However, few librarians have a solid statistical background. Statistics is not routinely taught in high schools, and is not a hot course in college. Few library preparation programs delve deeply into quantitative data and its analysis. Even master's programs tend to discuss statistics at the descriptive level (such as mean, mode, variance), and are most likely to work with surveys and numerical data generated automatically by library integrated management systems (such as holdings and circulation figures) and subscription database aggregators (such as access and download numbers for journal titles). Fortunately, librarians do not have to be statistical wizards, and when conducting data analysis should seriously consider getting help from local statistics experts in higher education (including graduate students). On the other hand, librarians should be competent consumers of research and statistics in order to make sense of data, and optimize its usefulness.

BENEFITS OF QUANTITATIVE DATA ANALYSIS

The benefits of quantitative data analysis for library improvement almost always result in a good return on investment. The systematic approach helps to prioritize and target efforts, thus optimizing time management. Data analysis can result in the most bang for their buck because performance and productivity improve, and quality is controlled, so not only are expenses reduced, but satisfaction by staff and users increases.

This sample case study demonstrates the power of quantitative data analysis. A university library wanted to deal with staff budget reductions and resource allocations as it impacted student service desk hours. Specifically, library administration wanted to reduce scheduling time, reduce schedule change, reduce errors, and improve student morale. Student staff scheduling, which was done by a non-student permanent staff member, was very labor-intensive. Some of the issues included need for flexible work schedules because of academic demands, and specific skills needed at specific times. It was determined that the students should do their own scheduling, but there were concerns about lowered productivity. To analyze the data, library administrators used a cause-and-effect matrix that weighed the effect of combined inputs and outputs based on relative importance. They then created a failure model and effect analysis to calculate a risk priority number (RPN) for each input/output process: SEV (severity of effect to customer), OCC (failure frequency), DET (how well failure is detected): SEV × OCC × DET = RPN. The

www.alastore.ala.org

highest RPNs emerged for emailing schedules for review, second highest RPNs were associated with schedule jockeying, and other risk priorities included knowledge of work requirements, human error, competing interest, emailing for student availability, and rehiring marginal students. Based on the analysis, library administrators created three levels of expertise, made a scheduling process template with a set skill level (including a new student supervisory level) and student number for each hour, determined the number of hours students could sign up for, and let high-performing level 2 and 3s sign up first. As a result of the changed scheduling process time involved in doing scheduling decreased, processes were streamlined, and human errors disappeared. The process also led to clearer expectations as well as closer supervision and assessment (Jankowski, 2013).

WHAT THIS BOOK PROVIDES

This book is intended to serve as a practical introduction to data analytics as a means for library improvement.

- Part I concludes with a basic model for library improvement, Six Sigma, and its variations are explained in chapter two.
- Part II includes five chapters, each of which details of the Six Sigma steps for improving library operations and customer satisfaction.
- Part III deals with data. Chapter eight explains how to clean data, chapter nine discusses how to match data with appropriate data analysis techniques, and chapter ten serves as a statistics primer.
- Part IV consists of 14 case studies that exemplify different library functions and associated data analysis.

With this book in hand, librarians can venture into the world of data, and leverage its use for informed, effective library improvement that positively impacts its stakeholders.

REFERENCES

Jankowski, J. (2013). Successful implementation of Six Sigma to schedule student staffing for circulation service. *Journal of Access Services, 10*(4), 197–216.

Soria, K. Fransen, J., & Nackerud, S. (2014). Stacks, serials, search engines, and students' success: first-year undergraduate students' library use, academic achievement, and retention. *Journal of Academic Librarianship, 40*(1), 84–91.

www.alastore.ala.org

2

Planning with Six Sigma

Data analysis might be an interesting intellectual endeavor in itself, but its power lies in its leverage to take effective action. As such, data analysis comprises one aspect of library planning and management. Not that data analysis is a stand-alone step. Deciding which data to collect, how to collect data, and what to decide as a result of the data analysis all impact the success of data analysis.

Several planning models incorporate data analysis. However, for targeted library improvement, one of the most effective data-driven models is Six Sigma. This model focuses on processes, which provide a rigorous foundation for library program direction and implementation. Although Six Sigma is known more in business circles, it works well in the library environment. However, it is not a management panacea; this chapter lists critical conditions for its appropriateness and success, and ways to adapt some of its practices that fit local circumstances.

www.alastore.ala.org

HISTORICAL DATA-DRIVEN MANAGEMENT MODELS

As libraries have grown in size and complexity, the need for systematic management approaches has also grown. In general, library administration has drawn from other disciplines such as business and public policy. Data-driven administration probably has its roots in scientific management, which focused on empirical evidence to increase worker efficiency. This systemization of monitoring led to ideas of quality control and operations research.

However, data-based management also extends to the idea of knowledge management: systematically using information to make appropriate and timely adjustments so that organizations can respond to internal and external changes and act strategically. The goal continues to be optimal organizational practice and improvement.

In the 1980s, total quality management (TQM) became popular in the business sector. The underlying idea was to institute an organization-wide climate of continuous improvement to deliver high-quality products and services cost-effectively. Top management has ultimate responsibility for quality improvement, which includes not only systematically monitoring processes but also leading and supporting all the employees. Quality itself is ultimately defined in terms of the target clientele: their needs and wants.

One aspect of TQM is statistical process control: measuring processes to optimize quality and minimize variation (quality control) (Wheeler & Chambers, 2010). To that end, the variables of each process need to be identified and measured, such as the supplies (quality, cost) and their interface or combination (such as attaching a spine label on the book); labor (performance quality and quantity, salary) and their interaction (e.g., one step to another, collaboration, etc.); and the work environment (e.g., furniture, traffic flow, lighting and air quality, etc.). Each variable is analyzed and controlled so that the entire process can be optimized. Each process can be considered as a micro-system, contributing to the overall system of the organization.

Several current management models focus on organizational improvement through a cycle of planning, implementation, study, and action that focuses on data analysis.

After World War II, the US Navy applied statistical process control to its own operations, which formed the basis of the US Department of Commerce's Baldrige Performance Excellence Program. The Baldrige Excellent Framework helps organizations assess their performance management system by measuring and analyzing the effectiveness of their leadership, workforce, operations, strategy, and customs to attain optimum results.

Largely supplanting TQM, ISO 9000 consists of a collection of standards. ISO has an international trade focus, and examines the links between

www.alastore.ala.org

suppliers and insurers. ISO 9000 is based on eight management principles that insure that organizations meet legal product requirements.

Kaizen, which means "continuous improvement," started after World War II in Japanese businesses. Its philosophy is that improving standardized activities should involve all employees in identifying and eliminating waste in all organizational functions. Kaizen promotes employee-initiated small scientific experiments to make and monitor change rather than top-down large projects. Kaizen also incorporates quality circles in which employee teams take responsibility for suggesting and making changes. The steps involved in Kaizen include: standardizing operations, measuring, comparing, innovating, and upgrading the standardization.

The Lean approach focuses on cutting out waste, and using only value-added process steps. The ultimate measure is customer satisfaction: Is the person willing to pay for the product or service? The main steps in the Lean approach include: analyzing opportunity, planning improvement, focusing on improvement, and improving and delivering performance.

SIX SIGMA

Six Sigma is the most well known of these data-driven models for organizational continuous improvement. The business sector in particular uses Six Sigma as a management tool to optimize cost-effective practices, control quality, and increase customer satisfaction by using data to identify problems and their causes, and then identify workable solutions. As such, Six Sigma has several key attributes: customer focus, data driven, rigorous process improvement methods and tools, professional development, full time resources, strategy execution, and quantifiable results (Galganski & Thompson, 2008).

Six Sigma was conceived by Bill Smith, a Motorola Corporation reliability engineer. He realized that complex organizations include many processes, each of which might have a sizable failure rate. Because processes can impact each other throughout the system, the rates could multiply and the resultant failure rate could end up being very significant. Smith asserted that to maintain quality control, any one process should have an error or failure rate that is six standard deviations from the norm: less than 3.4 defects in a million units. When such high standards are met, productivity and profits increase, and clientele are likely to value the organization more. Because standard deviation is symbolized mathematically as a small sigma, σ, the term Six Sigma was coined to capture the essence of improvement goals in this approach.

www.alastore.ala.org

The management principle behind Six Sigma is one of data-based enterprise-wide involvement and resource allocation. As such, Six Sigma includes two main models: DMAIC (Define, Measure, Analyze, Improve, Control), which is used for improving existing processes; and DMADV (Define, Measure, Analyze, Design, Verify), which is used for developing new processes.

DMAIC

The following steps detail DMAIC (Brassard et al., 2002).

1. *Define the project.* The organization's stakeholders identify a project that is likely to result in significant improvement relative to the resources required to get those results. The group clarifies the project's purpose, scope, and value. Then the group documents background information about the process under scrutiny and the impacted clientele. With these data, the group can identify which resources are required and available. Taking into account the clientele's expectations, the group then determines the key characteristics by which to measure success. Next, the group develops a written organizational agreement and a communication plan, including an intended timeline. With the high-level process outlined, the implementers are chosen: leaders, managers, and impacted staff.
2. *Measure the current situation and performance.* At this point, the project team investigated the identified problem in detail: what the problem is or where it is happening. This process requires collecting baseline data, so the team has to determine which data are essential to collect in order to identify the defects/problems and their possible causes; what are the key performance indicators? The team also has to decide which tools to use to collect those data. With this information, the team can see how the current process operates, and decide on targeted improvement performance level, which becomes a focused problem statement.
3. *Analyze the problem's root causes, and collect evidence to support that stance.* Using the collected data, the team applies statistical methods to find cause-effect relationships between processes and results. Statistics enables the team to test multiple causes of the program, such the specific patterns found relative to materials, human actions, and working environment. The team also determines the process's capacity to accomplish its task (i.e., are there enough staff to conduct a library orientation for every freshman class). Once the team comes to agreement on the cause of the problem, they can address that cause by hypothesizing feasible solutions.

www.alastore.ala.org

4. *Improve performance.* At this point, the team plans and tests interventions that can address the problem's root causes. Before-and-after data for each intervention are collected and analyze the compare performance. In some cases, an intervention makes no difference, which means another solution must be found. In terms of management skills, preparing staff for change must also be planned at this step.
5. *Control the process by standardizing practice.* The team seeks to maintain consistent high-quality performance. To this end, the improved practice is documented, and affected staff are trained to implement that process competently and consistently. Performance needs to be closely monitored at this time to identify and rectify deviations. The team also creates a process for updating procedures and anticipating future improvements. They also review efforts, and share lessons learned.

DMADV

DMADV follows similar steps. However, the chosen project focuses on a new initiative that needs to be developed, rather than improving an existing one. At the analysis step, the organization looks as the process options. At the design step, instead of an improvement step, the determined process is designed in detail. Then, instead of a control step, a verify step checks the design process and its ability to meet the identified needs.

Arizona State University's interlibrary loan process exemplifies the DMIAC model and its benefits (Voyles, Dols, & Knight, 2009).

Goal: To embark on a strategic project to assess the service quality and cost of filling interlibrary loan journal article borrowing requests.

Define: The Document Delivery Team's goal in this phase was to gain a clear understanding of the interlibrary loan (ILL) process, in order to improve its turnaround rate and cost-effectiveness for periodical articles. Their premise was that no additional staff or money would be required to meet their goal.

Measure: The Document Delivery Team created a flowchart of the ILL process; gathered ILL activity sheets; and set up a performance matrix to identify inputs and outputs, key success factors, quality standards, and cost for each step. The team interviewed the associated staff about the ILL process.

Analyze: The team's goal in this phase was to make sense of all the data gathered in the measure phase. The team created a fishbone diagram of contributing factors. They also produced a histogram that

www.alastore.ala.org

visualized the percentage of requests not filled in three days, which uncovered the program with weekend request fills. A subsequent Pareto chart verified that weekend lending was the root cause of lag time. They found that student workers executed the weekend ILL process because no permanent staff were present then. Drilling down, data revealed that student workers had difficulty choosing the right ISSN. Furthermore, difficult requests were set aside, a situation mainly experienced by the student workers.

Improve: The team recommended and tested the following solutions: having other evening and weekend permanent staff do ILL during their downtime, adding permanent staff hours on evening and weekends, training all relevant staff on ILL procedures, replacing student workers with full-time temporary staff, adjusting scheduling, and encouraging other libraries to increase their own evening and weekend ILL staff and use the consortia's union catalog more often.

Control: A follow-up XmR control chart highlighted data point patterns, which revealed decreased turnaround time after staff were trained and permanent staff were added to the weekend schedule. The control chart continued to be used to monitor ILL performance to insure sustained quality.

Results: The Six Sigma DMAIC approach resulted in an improved interlibrary loan borrowing journal article process. The modified process showed a cost saving of $2.09 per request, even with a 16% increase in borrowing requests. Furthermore, all interlibrary loan article deliveries were changed to be done electronically to clienteles' desktops, and turnaround time for article delivery was reduced to 70% of filled article requests being delivered in three days or less.

WHEN TO USE SIX SIGMA

No one program improvement process fits every situation, and Six Sigma is no different. When fully implemented, Six Sigma requires leader commitment to the project and to the dependence on data to make decisions. When Six Sigma teams have the capacity and support for innovative thinking, and the entire enterprise is involved, then results are likely to be more fruitful. People who use Six Sigma techniques should know about statistical process control techniques, data analysis methods, and project management. In its full manifestation, Six Sigma organizations incorporate specific, aligned training and certification for different project roles, and typically involve a corporate cultural change. However, Six Sigma's basic steps and many of its data collection

www.alastore.ala.org

and analysis tools may be applied to a variety of settings without subscribing to the entire package or using advanced statistics. Indeed, sometimes the library's capacity may be too limited to use certain data analysis methods. On the other hand, larger library systems can enlist the help of statisticians, such as university faculty, to help with some of the more sophisticated data analysis methods.

In the final analysis, cost-effective library improvement requires a systematic approach. Data are needed to describe and document current conditions, and facilitate analysis and decision making. Data provided the foundation for identifying reasonable interventions and measuring their impact. Elements of Six Sigma provide a workable framework for organizing efforts to problem solve efficiently and optimize improvement.

REFERENCES

Brassard, M., Finn, L., Ginn, D., & Ritter, D. (2002). *The Six Sigma memory jogger II*. Salem, NH: GOAL/QPC.

Galganski, C., & Thompson, J. (2008). Six Sigma: An overview and hospital library experience. *Journal of Hospital Librarianship, 8*(2), 133–144.

Voyles, J. F., Dols, L., & Knight, E. (2009). Interlibrary loan meets Six Sigma: The University of Arizona Library's success applying process improvement. *Journal of Interlibrary Loan, Document Delivery and Electronic Reserves, 19*(1), 75–94.

Wheeler, D., & Chambers, D. (2010). *Understanding statistical process control* (3rd ed.). Knoxville, TN: SPC Press.

www.alastore.ala.org

Index

www.alastore.ala.org

D

www.alastore.ala.org

www.alastore.ala.org

www.alastore.ala.org

www.alastore.ala.org

www.alastore.ala.org

S

www.alastore.ala.org

www.alastore.ala.org