The Search for New Measures: The ARL LibQUAL+ Project—A Preliminary Report

Colleen Cook, Fred Heath, Bruce Thompson, Russel Thompson

At its October meeting last year, the Association of Research Libraries launched a pilot project to assess service quality among research libraries. The project drew its impetus from the New Measures initiative of ARL, a joint undertaking of the Statistics and Measurement Committee and the Research Library Leadership and Development Committee. At a meeting in Tucson the winter of 1999, a gathering of library leaders from across North American research libraries articulated the need to develop alternatives to expenditure metrics as measures of library performance. While the descriptive statistics that ARL has developed over the decades remain the most important, and virtually only metric of assessment—they help identify those universities that have followed a practice of sustained investment over time and they document improvements and declines in resource allocation among member libraries—there is no necessary correlation between expenditures and service quality. In fact, the relationship that emerges between expenditures and institutional excellence is widely assumed but not empirically proven. The reality of such expenditure-based metrics is that rising in the ARL rankings only requires spending more.

At the New Measures conference, participants recognized the focus on expenditures to be widely at variance with new demands for evaluation and accountability. The pilot project proposed and endorsed at the October annual meeting, derived from SERVQUAL (for SERVice QUALity), addresses user assessments of service delivery. By measuring the relationship between service delivery and user satisfaction, librarians hope to control costs by directing resources to those service quality issues identified by users as most important and most in need of attention.

The pilot project reexamines traditional methods of assessing effectiveness while testing new theories to measure delivery of services to the consumers of research library resources. It is vital to recognize that research libraries are not alone in this predicament. The libraries at all postsecondary institutions face the same dilemma. Community colleges, four year liberal arts colleges, and the many other colleges and universities supporting graduate study must also rely upon expenditure metrics to assess their


<http://muse.jhu.edu/demo/pla/1.1/cook.pdf>
libraries. Thus, the need for new measures is common to all libraries. While the project’s research is grounded in the research library community, the emergent tools, with further research in other library cohorts, could be replicated at postsecondary libraries generally.

In the ARL-endorsed design, the LibQUAL+ instrument (an enhanced derivative of SERVQUAL) is being tested for score validity in comparison across libraries and across contexts, freeing institutions from reliance upon locally developed assessments. The project proposes to establish a national perspective, providing local institutions with data fine enough to diagnose local management issues. Where there are deficits, libraries will have the opportunity to make improvements that fit the local situation. A cohort of best practices across all the dimensions that define library quality may emerge, facilitating the efforts of administrators to tailor available resources to the institutional mission. Trends across the dimensions can be identified at the national level, placing local results in an important context for librarians and campus administrators alike.

Setting

At the ARL annual meeting, investigators asked for volunteers to participate in the pilot project, and thirty institutions responded. The majority of the costs for the pilot program was borne by Texas A&M University, where six years of experience with the administration of SERVQUAL provided the foundation for the study. However, each of the pilot libraries agreed to underwrite $2,000 of the costs, payable upon receipt of deliverables. We set an ambitious timeline for the project, calling for administration of the survey in April 2000 and delivery of the preliminary results at the July annual meeting of the American Library Association in Chicago.

Between October and April, Texas A&M administrators implemented a series of steps in fast-track fashion. From the institutions volunteering for the pilot, we selected a diverse group of twelve participants, and a planning session was scheduled for ALA Mid-Winter in San Antonio. While that process was unfolding, the design team undertook to reground the SERVQUAL instrument, already time-tested in a range of for profit and public sector environments over the previous decade. A series of interviews with library user representatives (i.e., faculty, graduate students and undergraduates) at the participating pilot institutions took place over the winter months. We also began work on survey design. In order to lend credibility to the project, the Cognition and Information Technologies Laboratory (CITL) at Texas A&M worked with liaison designates from each of the institutions to introduce the survey on each campus with a customized front-end web page. Completing the logistical requirements prerequisite for launch was the acquisition of the hardware and software required to administer the survey, capture the data, and analyze the results for a large-scale web-based survey spanning the continent.

The Institutions

Intending to capture a broadbased cadre of participants for the first pilot, the design team worked with the volunteers to ensure a wide range of ARL typologies. We preferred large public universities and private institutions, as well as those with urban missions,
land grant, and statewide responsibilities. It was important to ensure the participation of all regions of the United States as well as Canada. Making these choices was one of the more difficult early steps. But with the recognition that at least one more pilot phase would be required before settling on a final design, a number of institutions agreed to wait for the second iteration, and the following pilot participants were named:

- University of Arizona
- University of California, Santa Barbara
- University of Connecticut
- University of Houston
- University of Kansas
- Michigan State University
- University of Minnesota
- University of Pennsylvania
- University of Pittsburgh
- Virginia Tech
- University of Washington
- York University

From each university, the chief library administrator named a liaison to represent the library in planning activities with the Texas A&M design team. The liaisons and the library directors were invited to a planning conference during January Mid-Winter ALA meeting in San Antonio, where they discussed general design and timeline for the project.

At the meeting, CITL representatives presented the draft of the web survey instrument and discussed the processes by which the form would be customized for each campus. The timeline called for beta testing of the survey at the Texas A&M University Medical Sciences Library in early March. Because of the timing of the academic term, the final draft of the pilot survey would be launched at York University in late March, with the majority of the other libraries administering the survey in early April. The timing of its own academic calendar required that the University of Washington delay its survey until the end of the month. Texas A&M committed to collect the data and provide the pilot institutions with preliminary results of both local data and aggregate information in July at the ALA annual conference. The project's external quantitative evaluator led a discussion of the dimensions of service quality identified by the original SERVQUAL instrument in its decade of administration, and compared those to the dimensions recovered by Texas A&M in the library environment over the past six years.

Re-grounding of the Survey Instrument

Grounded in the Gap Theory of Service Quality, SERVQUAL was developed for the for-profit sector in the 1980's by the marketing research group of Parasuraman, Zeithaml and Berry (1985). Their ground-breaking research led to their development of the SERVQUAL instrument which undertakes to measure service quality across five dimensions:
• **Reliability**: i.e., ability to perform the promised service dependably and accurately;
• **Assurance**: i.e., knowledge and courtesy of employees and their ability to convey trust and confidence;
• **Empathy**: i.e., the caring, individualized attention the firm provides to its customers;
• **Responsiveness**: i.e., willingness to help customers and provide prompt service;
• **Tangibles**: i.e., appearance of physical facilities, equipment, personnel, and communications materials.

In the original SERVQUAL design, twenty-two questions are asked across the five dimensions the survey undertakes to measure. For each question, the customer is asked for impressions of service quality according to minimum service levels, desired service levels, and perceived performance. For each question, gap scores are calculated between minimum and perceived expectations and desired and perceived expectations. The zone of tolerance is the difference between the minimum and desired scores. Optimally, perceived performance assessments should fall comfortably within that zone. Administrators should be concerned with scores that fall outside the zone and falling trajectories over time, that while they may still reflect scores within the zone of tolerance, are nonetheless areas of concern. The following figure illustrates results (e.g., Q4 and Q9) falling outside the zone of tolerance.

![Radar Graph](image)

Figure 1. Radar Graph produced from SPSS data on Texas A&M Servqual
Building on the original model, the ARL project has its origins in the experiences derived at Texas A&M University Libraries over six years in translating the SERVQUAL instrument to the research library context. Administering SERVQUAL as an assessment tool for library performance in 1995, 1997, and 1999, the Texas A&M experience determined that the dimensions evaluated by the standard SERVQUAL instrument need to be adjusted for use in the research library context. Corroborating results found elsewhere in the literature, Texas A&M found only three library service dimensions isolated by SERVQUAL:

- **tangibles;** i.e., appearance of physical facilities, equipment, personnel, and communication materials;
- **reliability;** i.e., ability to perform the promised service dependably and accurately,
- **affect of library service,** which combines the more subjective aspects of library service, such as responsiveness, assurance, and empathy.

One of the first tasks at hand was to reground the instrument for the pilot project by visiting the pilot institutions and conducting a series of interviews with faculty, graduate students, and undergraduates in order to ascertain their views on what constituted quality library service, building theory, and revising the instrument during the survey period. Between January and March, the principal investigators visited many of the pilot institutions, and they conducted and transcribed a total of sixty interviews. While in-depth analysis of the qualitative results continues under the direction of the project’s external qualitative evaluator, preliminary results led to a revision of the survey instrument. In addition to the questions contained in the standard SERVQUAL instrument, the investigators added a battery of other questions to test for the potential of two additional dimensions identified during the interviews: access to collections and libraries as place—a concept transcending the definition of tangibles in the original SERVQUAL.

**The Sampling Frame**

One of the issues confronting the design team was how to structure the sampling frame from which to seek data on service quality. They considered four constituent groups in designing the sample: faculty, graduate students, and undergraduates—and library staff. Because of the importance research libraries place on satisfying the unique research needs of the first two groups, the investigators decided not to draw a proportional sample, which would possibly tilt the results in the direction of the identified needs of a larger undergraduate population. Rather, design efforts were directed toward a random sampling to produce roughly equal response sets for the three constituent groups. Because the theory underlying service quality presumes a full understanding of client needs, the instrument also undertook to measure how effectively library staff anticipated the perceptions of the constituent groups. As a rule, we asked the pilot libraries to draw from the campus e-mail address pool, a random sample of nine hundred undergraduates, six hundred graduates, and six hundred faculty. A single variation was permitted in one case where the total faculty population was only slightly larger than the targeted sample. The entire library staff in each case was encouraged to respond to the survey. The Texas A&M design and the survey plan on all campuses were submitted to and approved by human subjects review boards.
A criticism often levied at web-based surveys is the potential for sample bias that can arise when a large proportion of the targeted population is excluded from participating for one reason or another. It is difficult, for example, for a manufacturer of sporting goods to ensure a representative sample of consumers sufficient for reliable market analysis. The absence of a reliable and exhaustive source of e-mail addresses is one problem, access to the requisite computing equipment another. However, few such problems are encountered in the research university environment, where e-mail addresses are ubiquitous and access to the Internet via personal computers is commonplace. While the design team continues to evaluate the results for how representative they are and questions of nonresponse bias, the designers and participants felt that of all communities in North America, the research university community was best suited to respond to a web-based survey.

Infrastructure Issues

Overall, the design construct required that Texas A&M develop a system that was capable of handling the potential of twenty thousand survey responses in a very short period. In order to ensure as high a response rate as possible to a lengthy pilot instrument, considerable attention was paid to survey design. The goal was to develop an instrument that each respondent could answer as quickly as possible, preferably within thirteen minutes, while minimizing deflections prior to completion.

Texas A&M University procured and installed a Dell PowerEdge 4300 Server and two Dell Power Edge 2400 servers for the administration of the project. The two 2400 servers collected data from the potential respondents at the twelve participating institutions. The 4300 server housed the Microsoft SQL database software, capturing and channeling the result sets into SPSS for analysis. Members of the design team from the Cognition and Instructional Technologies Laboratory (CITL) at Texas A&M configured the servers and worked with the liaisons at the participating institutions to prepare their web pages and develop their samples. The survey instrument itself was written in Cold Fusion. Cold Fusion is a server side technology that allows dynamic generation of content from a database. Cold Fusion uses "html-like" tags inside standard web pages to integrate the web functions with a database to create dynamic web pages. Cold Fusion works with several web servers, and can access several different databases (including Oracle, Sybase, Microsoft SQL, Access) so that future applications of the system are transportable. SQL is the standard language for relational database management systems, is used by most common database systems, and is essential for the operation of the system.

After initial inhouse testing in the Texas A&M University main library, the web instrument was beta-tested with the Medical Sciences Library at Texas A&M in March 2000. The York version was loaded onto the Web March 15. Based on experiences there, the instrument was slightly revised for the April administration. The following figure illustrates a page from the web survey.
Please rate every item in all three columns by clicking the appropriate button to mark your rating. If the item does not apply to you then select the Not Applicable (N/A) check box.

<table>
<thead>
<tr>
<th>When it comes to...</th>
<th>My Minimum Service Level is (more info):</th>
<th>My Desired Service Level is (more info):</th>
<th>My Perception of the Library’s Service Performance is (more info):</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>low</td>
<td>high</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>1) Convenient access to library collections</td>
<td>c c c c c c c c c c</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>c c c c c c c c c c</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>2) Providing services as promised</td>
<td>c c c c c c c c c c</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>c c c c c c c c c c</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>3) Keeping users informed about when services will be performed</td>
<td>c c c c c c c c c c</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>c c c c c c c c c c</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>4) A place for reflection and creativity</td>
<td>c c c c c c c c c c</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>c c c c c c c c c c</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>5) Providing service at the promised time</td>
<td>c c c c c c c c c c</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>c c c c c c c c c c</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
</tbody>
</table>

This survey may be easier to see if you click the Maximize button on the top bar of this window.

Figure 2.

Prior to launch, a number of other administrative details were ironed out. To add to the survey’s credibility, a series of important steps were taken. First of all, with the cooperation of the Coalition for Networked Information (CNI), a DNS gateway was established between ARL and Texas A&M. When respondents at each university opened the URL given to them in the local e-mail message, they were transported to an instrument containing their institutional logo and bearing the Association of Research Libraries URL address. In fact, the survey was running on servers at Texas A&M in College Station. This approach modeled the ultimate goal of mounting the protocol on ARL servers. Secondly, each potential respondent in the sample received an e-mail message from their local library administrator explaining the nature and purpose of the survey and stressing its importance in the local context. In order to encourage completion, respondents were given the optional opportunity to register for a Palm Pilot drawing by placing their e-mail address in a lottery. Otherwise, the respondents were assured of complete respect for their privacy.

Preliminary Results and Future Plans

Descriptive Statistics

By the first week in June, the survey had run to completion on all campuses, and the design team began its analysis of the data. It is premature to share analysis of the data at this time. However, some useful preliminary information can be shared. Altogether, some five thousand responses to the survey were received. While analysis continues, it appears that an overall response rate of approximately 25 percent was achieved, within the range of large scale web-based surveys reported in the literature. Further, we appear to have avoided many of the experiences of earlier web-based surveys. It is commonplace to find that young males dominate the response set in such surveys. However, in this
instance the respondents were almost equally divided between male and female. Even more importantly, the survey obtained its goal of achieving in the aggregate equal sets of respondents from faculty, graduate students, and undergraduates, with age groupings showing a similarly satisfactory distribution. Large sets of respondents in the reporting disciplines were also achieved:

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business</td>
<td>406</td>
<td>8.2%</td>
</tr>
<tr>
<td>Education</td>
<td>308</td>
<td>6.2%</td>
</tr>
<tr>
<td>Engineering</td>
<td>506</td>
<td>10.2%</td>
</tr>
<tr>
<td>Fine Arts</td>
<td>199</td>
<td>4.0%</td>
</tr>
<tr>
<td>Health Sciences</td>
<td>497</td>
<td>10.0%</td>
</tr>
<tr>
<td>Humanities</td>
<td>529</td>
<td>10.6%</td>
</tr>
<tr>
<td>Law</td>
<td>93</td>
<td>1.9%</td>
</tr>
<tr>
<td>Library</td>
<td>588</td>
<td>11.8%</td>
</tr>
<tr>
<td>Science</td>
<td>859</td>
<td>17.2%</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>821</td>
<td>16.5%</td>
</tr>
<tr>
<td>Veterinary</td>
<td>109</td>
<td>2.2%</td>
</tr>
<tr>
<td>Other</td>
<td>65</td>
<td>1.3%</td>
</tr>
</tbody>
</table>

The possibilities of analyzing differences in assessments of library service quality by various types of users will occupy the remainder of the summer.

In the reporting out in July, each of the pilot libraries will be provided with mean scores on each of the questions as well as for each dimension the instrument succeeds in defining. Each participant will also receive the aggregate mean scores for both each question and each dimension, and other descriptive statistics as well. One of the many important milestones of the July session will be to assess the experiences of the pilot libraries in the administration of the survey on their home campuses. The design team has received the participant comments delivered to campus liaisons regarding the survey on their home campuses. The investigators are categorizing and subjecting those comments to content analysis. Significant issues range from the quality of the survey design, length and ease of completion, browser and operating system limitations, privacy concerns, and the like. Accommodation of these concerns is key to strengthening the survey instrument through its subsequent design phases.

**Next Steps**

Following the reporting-out session at ALA, a preliminary review of the findings will be presented at the sixty-sixth IFLA General Conference, Section on Statistics, in Jerusalem in August 2000. A comprehensive look at the results will be reported at an ARL international conference on the Culture of Assessment, in Washington, D.C., October 2000. Upon the conclusion of testing and assessment, the collaborators will issue a monograph assessing the cross-institutional data on each of the service dimensions.
The ARL-sponsored monograph will include the information on aspects of quality library service derived from the interviews at the twelve participating universities. It will also focus on the practical aspects of implementing and administering a large-scale survey across the Web. LibQUAL+ will be evaluated for its utility as a best practices tool for research libraries. Concurrent with the completion of the monograph, we will disseminate the findings of the first pilot project at the fourth Northumbria International Conference on Performance Measurement in Libraries and Information Services in 2001.

Over the summer of 2000, after all data have been collected, we will subject the theoretical foundations of the instrument to rigorous quantitative testing. These analyses will be grounded in the premise that scores, not tests, are reliable and valid. Thus, it cannot be assumed that just because SERVQUAL functions well in business settings that scores from the same protocol when used in library settings will also have sufficient psychometric integrity.

First, score reliability will be evaluated. These analyses will examine "corrected" item discrimination coefficients and alpha-if-deleted statistics, as well as total score alpha coefficients. Second, the primary methods for evaluating validity will invoke factor analysis. As J.C. Nunnally noted, "factor analysis is intimately involved with questions of validity... Factor analysis is at the heart of the measurement of psychological constructs." 6

In the academic year 2000-2001, the instrument will be further refined. From among the respondents of the first phase, some may be tagged for a longitudinal follow up study. In this manner, it will be possible to test the findings qualitatively by going back to some of the respondents in online focus groups. A number of libraries have already expressed interest in being included in the second pilot in the Spring of 2001. We expect that the number of participants double in the second phase, and other types of libraries may be included.

The third year will mark the emergence of a mature instrument and its movement from the design oversight of Texas A&M University to operational administration by the Association of Research Libraries. ARL may acquire equipment and software similar to that procured, configured, and developed by Texas A&M, for ongoing administration.

Summary

The strength of the project is the rigor of its design and the robustness of the statistical analysis to which the results will subjected. Close peer scrutiny of the findings is assured through broad dissemination of the results. The model recognizes the pre-eminence of local findings and surfaces best practices across institutions. If successful, the pilot project will be scaled to a national undertaking, accommodating other related research. The experience will enable a technology transfer to libraries generally as well as to a broad range of related ARL applications.

The project plan envisions the migration of the operational oversight of the tool to ARL by 2002, with the instrument available for widespread administration. The advantages of an assessment tool, well grounded in theory and rigorously administered, hold promise to answer finally the calls for greater accountability and responsiveness to user needs in college and university libraries.
Notes