

Proposal Title: Collaborative Assessment: North American Academic Libraries' Experiences Using the Balanced Scorecard to Measure Performance and Show Value.

Authors: **Mengel, Elizabeth**, Associate Director, Scholarly Resources and Special Collections, The Sheridan Libraries, Johns Hopkins University, emengel@jhu.edu. 3400 N. Charles Street, Baltimore, MD 21218 USA 410-516-0809.

Lewis, Vivian, Associate University Librarian, Organizational Development Mills Memorial Library, McMaster University, lewisvm@mcmaster.ca, 1280 Main Street West, Hamilton, Ontario L8S 4L6 Canada 905-525-9140, x23883.

Background

The Balanced Scorecard is a widely-accepted organizational performance model that ties strategy to measurable results in four critical areas: finance, learning and growth, customers, and internal processes. While originally designed for use in the for-profit sector, the scorecard has been adopted by non-profit, government, and other mission-based organizations including some academic libraries. This paper focuses on the continued experiences of two prominent North American research libraries, Johns Hopkins University and McMaster University. These two libraries were part of an Association of Research Libraries (ARL) pilot effort that included a total of four institutions, the two represented by the authors, plus the University of Virginia and the University of Washington. (For the purposes of this paper, the four libraries are described as Cohort 1.)

The ARL initiative was described as “an investment in helping libraries make a stronger case for the value they deliver by developing metrics that are tied to strategy.” [E-News for Library Directors, January 2009, <http://www.arl.org/news/enews-15jan09.shtml#32>]

In addition to developing cases for value, the ARL initiative aimed to test the viability of creating a community of practice to facilitate scorecard implementation and practice. As the original initiative progressed with Cohort 1 in 2009-2010 and strategy maps, objectives, and measures were developed, the participating libraries discovered common themes and threads running through their respective work. The libraries had very similar focuses in each of the four scorecard perspectives (financial, customer, internal process, and learning and growth). These commonalities intrigued the authors and encouraged further scrutiny.

Three key questions emerged. First, given the commonality in objectives, how much overlap would emerge in the actual measures chosen by the Cohort 1 libraries to track performance? Second, given the long-standing nature of the formal ARL statistics program, what portion of the measures that were developed for the scorecard draw from the annual statistics reported to ARL? Finally, as ARL continues to move the scorecard initiative forward and 10 new local sites (described in this paper as Cohort 2), join the program; can new tools be created to further the concept of collaborative scorecard development? Two approaches seemed worthy of exploration: a) the compilation of an inventory of all measures from which prospective scorecard sites could choose or b) create a “common” set of uniform core measures for use by scorecard participants.

A note on the time frame of the data used in this paper; while the four libraries in the original ARL initiative started at the same time, organizational acceptance, budgetary cycles, and strategic planning cycles are different across the institutions. To eliminate confusion, the authors will use naming

convention of YEAR 1 for the first cycle of work the libraries completed and YEAR 2 for the second cycle. These cycles roughly correspond to 2009-2010 for Year 1 and 2011 for Year 2.

Common Objectives

Within the context of the scorecard, objectives are the building blocks of a strategy. They define the organization's strategic intent and for mission-based organizations they connect directly to achieving its chosen mission. It is reasonable to assume that research libraries should have common objectives. The work research libraries are performing is very similar, the missions are often surprisingly similar, and the institutions research libraries support have similar purposes. For example, all research libraries need to secure funding, there are common concerns about physical and virtual space, there are significant changes happening in regard to staff recruitment and development, and there is a need to ensure that our intrinsic value is known by stakeholders and patrons.

Not surprisingly the Cohort 1 libraries' objectives displayed significant similarities. A detailed description of the common objectives can be found in Lewis' 2010 paper, "Building Scorecards in Academic Research Libraries: Organizational Issues and Measuring Performance." Given that the Cohort 1 libraries were all trying to achieve similar objectives, the authors were eager to explore the extent of overlap in the resultant measures. Would the libraries create similar measures for those objectives or would local concerns and needs force divergence?

Measures

In the Balanced Scorecard, measures tell an organization if they are making progress in implementing their strategy, and whether that strategy is actually having the anticipated outcome on the planned organizational performance. In other words, are we doing what we said we would do and does it affect our users, stakeholders, and staff as we intended. For the measures analysis portion of this paper the authors will only use measures from Johns Hopkins, McMaster, and the University of Washington libraries. The University of Virginia Libraries had been using the balanced scorecard since 2001 and the measures they had at the time of this paper were in the process of being refreshed.

Libraries have a long tradition of collecting statistical and other measures related to library performance. However, these measures on the whole tend to be input measures. Input measures traditionally count things such as the number of volumes, user seats, librarians, dollars, or computers. They form the basis of many of the regional or national statistical surveys where comparisons can be made. The authors used Wordle to examine the wording of measures developed by the three libraries. As illustrated by the word cloud, libraries still rely heavily on counting. The emphasis on sheer quantification raises critical questions. Are libraries counting the right things? Are libraries counting the things that matter to our patrons, to our stakeholders, and for our staff?

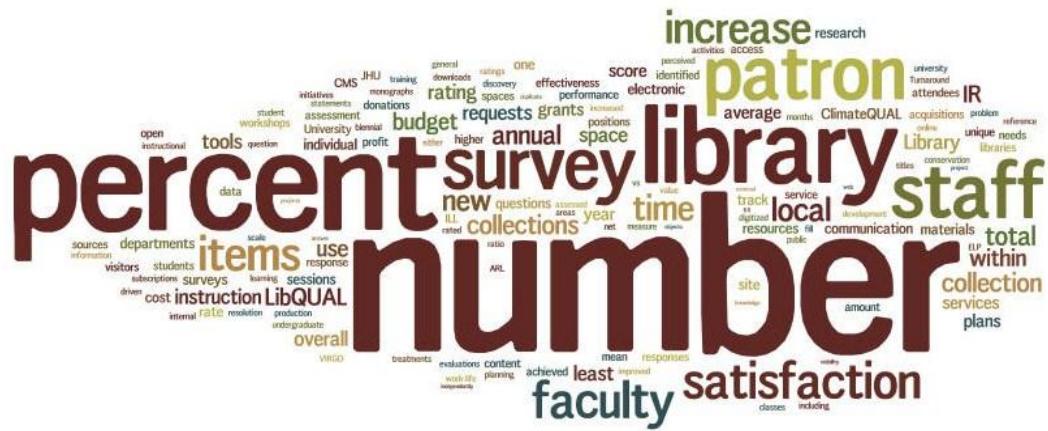


Table 1 indicates the distribution of Cohort 1 measures across the four scorecard perspectives. Given libraries' historic focus on patron services, it is not surprising to see the concentration of measures in the Customer perspective. The measures in the Customer perspective also tend to be less about counting how many of something and more about service effectiveness based on specific output-related measures.

Table 1. Year 1 Aggregated Measures

Table 1. Year 1 Aggregated Measures		
Perspective	Number of Measures	Typical Measures
Customer	41	Effectiveness of library instruction Effectiveness of liaison librarians and the services they provide Collections related activities
Financial	19	Material inflation Revenue generation including grant revenue ARL Investment Index Donations
Internal Processes	19	Collections and the use of those collections Patron satisfaction with collections Web and IT services Promotion of the library
Learning and Growth	15	Employee satisfaction Staff retention rates Staff ratings on various surveys including ClimateQUAL Numbers of assessments or development plans

The authors also examined the connection between the scorecard measures and ARL standard statistics. The measures developed by Cohort 1 libraries made very little use of current ARL statistics. Overall, only 9.5 percent of the measures were related to the ARL statistics. The most commonly used ARL statistic was ARL Supplemental Statistic number 10 which deals with library digitization activities including number and size of libraries digital collections, number of collections, size of collections, and number of items. Table 2 provides an overview of the ARL statistics used as scorecard measures.

Table 2. Measures Using ARL Standard Statistics	
Supplemental Statistics 10b	4
Supplemental Statistics 11a	2
Supplemental Statistics 10c	1
ARL Investment Index	1
Standard Statistics 5	1
Standard Statistics 16	1
Standard Statistics 35-36	1
NOTE: for information on ARL's statistics see http://www.arl.org/stats/annualsurveys/arlstats/	

Another common measure methodology observed was the use of surveys. Table 3 provides a distribution of surveys used as measures. LibQUAL was largely used in the customer perspective to identify both customer satisfaction and space related issues. ClimateQUAL was only used in the Learning and Growth perspective and related to library staff. The WOREP and READ scales were used as teaching and learning measures. Further investigation is needed to determine if it would be feasible to create collaborative work around the other locally created surveys.

Table 3. Year 1 Measures Utilizing Surveys	
Locally created surveys	15
LibQUAL	4
ClimateQual	2
WOREP (Wisconsin Ohio Reference Evaluation Program)	1
READ Scale (Reference Effort Assessment Data)	1

The question remains if libraries can improve their arsenal of assessment tools by working *alongside* each other (as opposed to directly *with* each other) as they implement local organizational performance measurement instruments? To answer that question the authors propose using the experiences of the second cohort group to explore options of measurement development.

The Benefits of a Collaborative Approach

As noted earlier, the ARL initiative aimed to test an alternative, more collaborative approach to assessment tool development. Each library aimed to create its own scorecard. Would local teams benefit from working alongside their counterparts in other ARL libraries, even if each team was producing an independent instrument?

Interviews with members of the Cohort 1 libraries point to very clear benefits from the collaborative approach. Participants described the opportunities for face-to-face and teleconference discussions as “invaluable” to the success of their local implementations. They frequently consulted the other participant’s scorecards for ideas. Three benefits were identified: 1) seeing what other institutions were doing (or considering doing) **saved time**. The ideas they took away were “kitchen tested” in another like organization; 2) the concepts had **high credibility**. Knowing that something had been done in another institution often came in handy when persuading reluctant colleagues to accept a new idea;

3) Finally, the idea of adopting a measure similar to someone else's laid the ground for future comparisons and **benchmarking**.

The process was not seamless. Cohort 1 team members described the frustrating experience of being initially persuaded by a specific measure in use at another institution – only to discover that the measure was either inactive or fraught with pitfalls. Reading published lists of measures failed to reflect the true value of various components.

The Cohort 1 also noted the challenges associated with getting the measures "right" or at least "good enough" for early consumption. By way of example, the McMaster team started with a cluster of 63 measures in support of their 13 objectives, far in excess of the 16-36 recommended measures. The initial slate included a host of ambitious but untested concepts. The final slate released a year later bore little similarity to the initial roster. In many cases, it wasn't until actual data was collected, massaged and analyzed that the Scorecard team knew they had captured a "measure that mattered". Johns Hopkins held onto a slate of 55 measures the first year, far too many for effective use. The second year the slate was reduced to 20 measures.

The refinement process was exhausting, especially for institutions with less fully-developed assessment programs. In many cases, the teams were exploring data sets for the first time. Understanding how large blocks of data could be boiled down to a single trackable scorecard measure proved particularly difficult. For example, most sites collect large amounts of data associated with their instruction program, but what single meaningful measure could be tracked over time? The librarian's natural reluctance to filter information often became an impediment.

Enhancing Collaboration

Once the call for a 2nd cohort of Scorecard sites was initiated, members of the Cohort 1 began discussing strategies for facilitating measure development: Could they use their own experiences over the previous several months to ease the load of the 2nd cohort sites and, in doing so, lay the groundwork for future, more robust scorecard implementations? After much discussion, two possible models emerged for furthering the collaboration agenda: a) the compilation of an inventory of all measures in use at Cohort 1 sites; and b) the crafting of a core set of common measures reflecting a synthesis of thought at all sites.

A The Inventory

The inventory is envisioned as a compilation of all measures in use at Cohort 1 sites. Each entry would include: measure name, formula, originating library, an indication of whether the measure was active or inactive at that site, and a rating (by the local site) of how successful the measure is in capturing meaningful data. For example:

Table 4. Inventory of Measures Example

#	Measure Name	Formula	Originating Library	Active / Inactive	Rating (0-5) (N/A = inactive)
1.2	Study Seats	Number of study seats located in	McMaster	Active	5

		“silent” or “quiet” areas.			
--	--	----------------------------	--	--	--

The authors recognize some definite pluses and minuses associated with the inventory model. Such a tool would compile options from which local sites could choose based on their environments. The inventory could be pulled together very quickly – without a lot of negotiations or compromises. The tool would also facilitate the sharing of specific information (e.g., the effectiveness rating). At the same time, the inventory model lacks the consistency required for benchmarking. Individual libraries could address a similar theme, but use a completely different approach and language to capture it.

B. The “Common Set”

The “Common Set” is envisioned as a tight set of measures carefully crafted by Cohort 1. The set would reflect, not only a common approach to measuring a particular theme, but also common language. Common measures might adopt the thinking of one Cohort library over the others or, alternatively, reflect a compromise statement not actually in use at any Cohort 1 site. For example, a common measure linked to staff satisfaction/healthy workplace could read:

Table 5. Common Set Example	
Measure Name	Formula
Healthy Work Place	Increase in the percent of staff answering “3” or above on ClimateQUAL question #

The uniform nature of the common set has both strengths and weaknesses. The consistent approach and language would facilitate benchmarking: Sites choosing to use the common measure could be assured that other adopting locations would use the same formula and language. The process of creating the set, however, would not be easy: Cohort 1 libraries would have to negotiate the measures between themselves. Additionally, libraries choosing to use the set would effectively give up the ability to customize language to meet their local environments.

Research Design

The authors decided to test the approaches as part of a small-scale research project. They will construct the two instruments in conjunction with Cohort 1. The two sets will be presented to Cohort 2 at a face-to-face meeting. Three months following this presentation, the authors will interview members of Cohort 2 regarding the perceived usefulness of the two sets for their local work. A second set of interviews will be scheduled for a few months later to determine actual measure adoption. Results will be presented in a formal report and presented at an upcoming assessment conference.

Conclusion

The Balanced Scorecard is a powerful tool for tracking success in meeting organizational mission. Implementing a scorecard effectively requires significant intellectual effort. In many ways, setting clear objectives and identifying crisp initiatives is the easy part. The challenging work lives in the middle – in the identification and fine-tuning of meaningful, well-aligned measures.

The current slate of measures in use by Cohort 1 shows a fair amount of similarity – especially in areas of Customers and Learning and Growth. Given the clear overlap in objectives, this level of measure convergence is not surprising.

The current slates illustrate some linkages to the ARL statistics and assessment program. The authors note that the use of annual survey data and large-scale initiatives like LibQUAL and ClimateQUAL is there, but not in the quantity one might expect. The small number of linkages suggests that further work needs to be done on the data being gathered by the program. (An ARL Statistics & Assessment Committee task group is currently working on this issue.)

The Cohort 1 experience with measure creation suggests that collaborative efforts pay off. The four teams benefitted greatly by working together: They saved time, leveraged the credibility of the larger group and laid the groundwork for future benchmarking. Further research will be done in the coming months on two instruments for furthering this collaboration. The authors will build two prototypes for discussion with Cohort 2 sites – an inventory of all used measures AND a set of common measures from which other libraries may opt to choose.

Selected References

Durgan, R. E., Heron, P. and Nitecki, D. A. (2009) *Viewing Library Metrics from Different Perspectives: Inputs, Outputs and Outcomes*. Santa Barbara, CA: Libraries Unlimited

Kaplan, R.S. and Norton, D.P. (1992) The Balanced Scorecard: measures that drive performance. *Harvard Business Review*, 70 (1), 71-79.

Kaplan, R.S. and Norton, D.P. (1996) *The Balanced Scorecard: translating strategy into action*. Boston, Harvard Business School Press.

Kaplan, R. S, & Norton, D. P. (2001). *The strategy-focused organization: how balanced scorecard companies thrive in the new business environment*. Boston, Mass.: Harvard Business School Press.

Kyrilidou, M. "Library Assessment as a Collaborative Enterprise" special issue of *Resource Sharing and Information Networks* on the theme "Creative Collaborations: Libraries Within Their Institutions and Beyond" 1/2 (2005/2006): 73-87

http://www.libqual.org/documents/admin/kyrilidou_haworth_sept72004.pdf

Kyrilidou, M. E-News for Library Directors, January 2009, <http://www.arl.org/news/enews/enews-15jan09.shtml#32>

Vivian Lewis et al, "Building Scorecards in Academic Research Libraries: Organizational Issues and Measuring Performance," ARL Library Assessment Conference Proceedings [forthcoming]

Self, J. (2004) Metrics and management: applying the results of the balanced scorecard. *Performance Measurement and Metrics*, 5(3), 101-105.