EDITORIAL

Moving from evaluation to assessment

DOI: http://dx.doi.org/10.3163/1536-5050.102.4.001

A few years ago, my health sciences library held an open house aimed at preceptors. The event was a big success: Many people attended, and feedback was positive. Yet when we tried to repeat the event in subsequent years, attendance foundered. We wondered, ''What had changed?'' Had most of the interested preceptors attended the first time? Did our advertising reach a different audience the second time around. We thought we had done exactly the same thing as before, but clearly something was different. Unfortunately, we had no means to identify what that difference was.

Ours was not an isolated experience. All too often, health sciences libraries launch a new class, adopt a new weeding method, or upgrade a website and report extensive, mostly positive, feedback. Student evaluations of the class are positive, the weeding project is completed in record time, or an online survey indicates that users like the new website. The librarians in charge of the project are happy, the director is happy, and the libraries may even present a poster on the project at the next Medical Library Association (MLA) annual meeting. Then others try to repeat the experience: they give the same class, weed a different collection with the same method, or use the same techniques on a different website. This time, it does not work so well. Or, perhaps, as in my case, the same library repeats the innovation but does not achieve the same results. Students in the new class are dissatisfied, the weeding takes longer than it should, or feedback suggests that the web revisions are confusing. What happened? It is hard to say.

Obviously, there are many variables that could account for these inconsistent results. The students who attended the first class may have been different from those in the second. The collection initially weeded so speedily might have been in nursing or internal medicine, while the second attempt focused on a different subject, with different usage patterns, such as biochemistry. The menu structure that worked so well was used for the original library's home page, and a second library tried to apply it to sub-pages of library guides. Alternatively, the reasons for the inconsistent results might lie elsewhere. Perhaps the first class was offered in the fall and the second in the spring, so student schedules were different; the data used for weeding at one time might have been higher quality than that available at another time; or maybe it was not the content of the second set of web pages but how faculty at the second institution used the library that made the difference.

These examples, although perhaps a bit forced, help to illustrate the difference between evaluation and assessment. In all these cases, the library evaluated its performance. The feedback it received suggested that what had occurred was successful, so, understandably, the library attempted to repeat the experience or perhaps colleagues tried to profit from it. However, the library did *not* do an assessment. There was no attempt to *assess* the reasons underlying the initial success. The results were positive, and the library assumed from that positive evaluation that the method was sound and should be repeated. However, the positive evaluation was merely a global indicator of success. It provided no information on what had led to the results. Thus, when the library tried to repeat the experience and failed, there was no way to determine why the results were not the same the second time around.

What is the difference between evaluation and assessment? In an instructional setting, evaluation refers to making a judgment or determination concerning the quality of a performance, work product, or use of skills against a set of standards [1]. Evaluations are designed to *document the level of achievement that has been attained*. Assessment, on the other hand, is focused on measuring a performance, work product, or skill in order to offer feedback on strengths and weaknesses and to *provide direction for improving future performance.* Assessments are nonjudgmental and are designed and intended to be used to produce improvement. In schools, students receive frequent evaluations in the form of grades on exams, classes, papers, and so on. These are global measures that tell the instructor whether the student has achieved a desired level of performance. Less commonly, students are assessed for gaps in skills, readiness to move to a higher grade level, language difficulties, psychiatric problems, and so on. These assessments help instructors to identify gaps and develop programs to address deficiencies.

Similarly, in a health sciences library setting, evaluation is a global measure of how well a particular project, class, or initiative performed and whether or not it met its goals. Assessment, on the other hand, helps the library to understand which elements of the project, class, and so on worked and where improvements could be made. In the open house I referred to initially, evaluation indicated that attendees enjoyed the event. An assessment would have provided information on whether the target audience had been reached, which elements of the publicity had been effective in reaching that audience, and what attributes of the event itself were critical to its success.

Whereas the educational establishment draws clear distinctions between evaluation and assessment, the library and information science field often uses the terms interchangeably. For example, the American Research Libraries (ARL), reporting on the results of one of their periodic ''Spec Kit'' surveys of academic library practices, states,

''To assess, in general, is to determine the importance, size, or value of; to evaluate'' [2]. According to that ARL report, academic library assessment activities typically begin with the desire to know more about a library's customers; almost half the libraries begin assessment with a user survey. The top five assessment methods at the time of the ARL report (2007) were statistics gathering, a suggestion box, web usability testing, user interface usability, and surveys like LibQual. Obviously, there is substantial value in understanding how people use the library, whether a website is usable, and which electronic journals are being downloaded; however, most of these techniques are aimed at evaluating performance rather than assessing the reasons why performance is succeeding or failing. Although the ARL report labels these as assessments, the education literature would refer to them as evaluations, because in most cases these tools tell library staff and administration the level of achievement that has been obtained, rather than provide directions for future performance.

In cases like these, a library may decide to undertake an actual assessment, designed to improve performance, at a later date. Having determined from a user survey, for example, that faculty are dissatisfied with the library's electronic journals, the library might conduct focus groups to identify the reasons for that dissatisfaction. After hearing from students that they value study space in the library, the library might convert some book stacks to a study area and then document the hours of use of the new facility. Learning from a usability study that faculty could not find information on proxy servers, librarians might compare their libraries' layout to best practices on website design to determine the source of the difficulty. Thus, in ''real life,'' these two concepts, evaluation and assessment, may merge into a single effort to gather data that will allow libraries to improve their performance.

The distinction between evaluation and assessment is critical to the *Journal of the Medical Library Association (JMLA).* One of the MLA goals, the goal that the *JMLA* is designed to support, is to:

> develop and manage a knowledgebase of health information research. That knowledgebase will enhance and improve professional practice, promulgate advances in biomedical research, demonstrate the value of health information for improved health, and support lifelong learning. The association will…promote the creation of knowledge through health information research and the practice of evidence-based librarianship. [3]

Creating this knowledgebase of health information research requires more than mere evaluation. While evaluation may indicate that a project met its goals, assessment is needed to explain exactly what worked and why, thus allowing health sciences librarians to advance and improve their performance. If the activities that libraries undertook were simple, this would not be the case. However, as in the examples at the outset of this editorial, libraries provide services in a complex environment in which many variables can, and do, affect performance. Thus to improve service delivery, librarians need to not only document that something worked (evaluation) but why it worked (assessment). For a student in a classroom, a ''B'' on a paper is not very helpful. What helps a student to improve performance is comments on a paper that suggest, for example, that the introduction could have benefited from reference to previous work on the subject. For libraries, hearing that another institution provided a well-received online class is not very informative; learning that introducing an online component led to an increase in retention is far more helpful.

Obviously there are reasons why libraries often choose to evaluate rather than assess their performance. Evaluations are usually easier to obtain. They may rely on only a single measure of performance, such as a course grade or a class rating. Evaluation instruments are also often easier to design. Questions on evaluation instruments frequently employ Likert scales that rate user satisfaction, while assessment instruments must break down the components of a service and measure the effectiveness of each one. For example, an evaluation of a class could ask students to rate the performance of the instructor, the setting, and their learning. This evaluation might also ask some questions that would measure actual skills acquired. An assessment of whether a new online class segment improved learning and/or user satisfaction would need, at a minimum, to ask questions that would allow comparison of learning and satisfaction in the online and traditional settings. A more rigorous assessment would try to identify what was responsible for reactions to the online component. Was it the ability to take the class at their own pace or the use of online tutorials or quizzes at each step providing feedback for learners?

Like many desirable outcomes, assessment is most easily achieved if it is planned for in advance. Before launching a project, time needs to be invested to consider what kinds of data will be required to assess it. To assess a weeding project that uses a new kind of data to improve the process, a library would need to collect information that would either allow comparison of time spent making decisions using that data to time spent in previous similar weeding efforts or collect data on the number of times using the new data made a difference in what material was weeded. To assess the effect of closing a service point, a library would need to collect data on use of other service points before and after, or on the number of complaints received before and after the closure. To understand the true success of my library's preceptor open house at attracting its target clientele, we should have asked for some demographics on the attendees. Note, however, that assessment does not

have to be quantitative. A library could just as easily interview patrons before and after the closure about how they were adapting to the change in services or interview staff members after a weeding project to determine if the new approach had made a difference. Sometimes qualitative data are far more revealing and helpful than quantitative reports when it comes to understanding how to improve performance in the future.

There is a famous quotation, usually attributed to the mathematician Piet Hein, which comes to mind here. ''Problems worthy of attack, prove their worth by fighting back.'' Assessment is not easy. Nevertheless, if, as health sciences librarians, we wish to improve the knowledgebase underlying our practice, we need to attack it. At the very least, performing more assessment, as opposed to mere evaluation, should be cost effective. Although it would take more time at the outset, it should result in more effective performance in the long run. Understanding what does and does not work and why will allow us to improve services to our users, satisfy our administrators, and bring quality health information to a wider audience.

Susan Starr, MLS, PhD, Editor-in-Chief, jmlaeditorbox@gmail.com, La Jolla, CA

References

1. Center for School Success. Assessment vs. evaluation [Internet]. West Lebanon, NH: The Center; 2011 [cited 5 May 2014]. <http://www.centerforschoolsuccess.org/index.php?option=com_content&view=article&id=86&Itemid=150>.
2. Association of Research Libraries. Library assessment [Internet]. Spec kit 303. Washington, DC: The Association; 2007 [cited 1 May 2014]. <http://arlstatistics.org/documents/admin/2013/2007_SPEC_Kit_303_Library_Assessment.pdf>.
3. Medical Library Association. Strategic plan of the Medical Library Association [Internet]. Chicago, IL: The Association; 2014. [cited 14 May 2014]. <http://www.mlanet.org/about/organization/strategic.html>.

STOP SEARCHING.

THE HEALTH SCIENCES EBOOKS YOUR PATRONS WANT

EASILY INTEGRATED WITH YOUR INSTITUTION'S WORKFLOW

START FINDING.

R2 DIGITAL LIBRARY

50 STM publishers,
1 web-based platform

Select the eBooks you need from thousands of best-selling current resources

Intuitive interface optimized for health sciences content

IF YOU HAVEN'T SEEN US LATELY, VISIT R2LIBRARY.COM AND START YOUR 30-DAY FREE TRIAL.

Rittenhouse Book Distributors, Inc.
511 Feheley Drive, King of Prussia, PA 19406