

# Meta-Assessment of Online Research Guides Usage

Charles H. Smith

**ABSTRACT.** A practicable approach to meta-assessment is reported involving the use of online research guides as an outreach tool at Western Kentucky University. A series of such guides has been produced that cover each department on campus; this approach lends its way to assessment efforts by identifying discrete and quantifiable target populations across campus. A multiple-regression model is used to predict usage rates on the basis of surrogates for size of potential usership, and variations in quality and size of the guides themselves. The “badness of fit” of the regression model can be used to suggest ways of improving the outreach effort, as the residuals left from the analysis arguably identify related lapses in coverage or promotion. If carried out on an annual basis, the approach provides an almost ideal services assessment tool, since various measures can be taken to improve those guides that during each period are identified as “underachievers.” doi:10.1300/J120v47n97\_08 [Article copies available for a fee from The Haworth Document Delivery Service: 1-800-HAWORTH. E-mail address: <docdelivery@haworthpress.com> Website: <<http://www.HaworthPress.com>> © 2007 by The Haworth Press, Inc. All rights reserved.]

**KEYWORDS.** Assessment, meta-assessment, outreach, reference services, multiple regression, stepwise regression, statistical analysis

## INTRODUCTION

All libraries collect various kinds of statistics that pertain to their operations and services; these are not only necessary for keeping track of day to day operations, but also provide means whereby institutions sup-

---

Charles H. Smith is Professor and Science Librarian, Department of Library Public Services, Western Kentucky University, Bowling Green, KY 42101 (E-mail: [charles.smith@wku.edu](mailto:charles.smith@wku.edu)).

The Reference Librarian, Vol. 47(1) (#97) 2007  
Available online at <http://ref.haworthpress.com>  
© 2007 by The Haworth Press, Inc. All rights reserved.  
doi:10.1300/J120v47n97\_08

porting the library can assess the continuing quality of their services. In reality, however, there are precious few statistics available that can tell us more about our operations than whether this or that flow has increased or decreased, or whether patrons seem to more or less approve of our current level of service. It seems that most statistical compilations—gate counts, interlibrary loan totals, and patron contacts, for example—in fact tell us more about ourselves than they do the patrons we are trying to serve. Most importantly, they relay but little information as to whether we are attending to all elements of our potential user base equally well, perhaps the most basic obligation we have related to fair and optimum allocation of materials and services.

Neither do such simple statistics well lend themselves to the ever-increasing need for assessment vehicles. Yes, we can use particular existing rates of service as an excuse for introducing some new contact point or collection, but in so doing we are merely guessing at both ends: What, objectively, can be the statistical criteria for recognizing relative current lack of service, and even worse, if we make some change to effect an improvement? Can we do more than simply guess what level of change will signify real betterment? Are we sure that a mere increase in usage, access, etc., truly corresponds to actual qualitative improvement?

In theory, this set of problems can be addressed by setting up means of monitoring the patron-library relationship that are recursive in nature—that is, those that both explicitly identify instances where service is, relatively at least, lacking, and provide feedback on what specific means may be employed to rectify the situation. This kind of “closing the loop” (Nichols and Nichols 2000) inherently produces information that lends itself to assessment and planning practices. As Peters (2000) has put it:

In addition to assessing the value of what is being assessed, the assessment activity should focus on the context and value of the assessment activity itself. Meta-assessment efforts also can examine the basic methods of an assessment program. For example, computerized monitoring (i.e., the use of computers to monitor human-computer interaction) in general, and transaction log analysis in particular, are interesting in the context of evaluating digital library services because they represent the possibility of making a service or online information environment automatically customized or self-improving. The system can contain within its programming the seeds of a self-evaluation. (The “self” here is the human-computer interaction as a series of events, or the online environment itself, not the human self.) A self-improving environ-

ment utilizes actual use of the system or service as data for making decisions and design changes that ostensibly improve the service for present and future users, both collectively and individually. Buttenfield (1999) states: "Ideally, one would prefer a self-evaluating and self-modifying system."

It is the purpose of this work to describe such meta-assessment system as it has operated at Western Kentucky University, Bowling Green, Kentucky. The history of the development of the system will be described first, followed by a detailed look at the monitoring process and its statistical analysis. The paper concludes with a brief consideration of the system's relevance to reference service assessment, including its relation to what we have termed "reference outreach" (Smith et al. 2000; Carson and Smith 2001).

### ***AN ONLINE RESEARCH GUIDES PROJECT***

In the fall of 1999 a task force set up within the Department of Library Public Services here met to consider some ways of improving our reference services program. One of my recommendations as leader of the task force was to initiate a system of Web-based guides that might aid the online user's investigation of various kinds of information sources, for example, our subscribed-to databases, and reliable services accessible through the free Web. The arrangement we decided upon consisted of two series of online "research guides."

The main set of guides (Departmental Research Guides: hereafter DRG) was to follow a basically subject-oriented organization, but differed from the typical arrangement by having each guide cover not a specific subject per se, but instead a specific academic department on campus. This series was to cover all departments, and have a uniform format across all guides, making it easier for users to move from one to the next with a minimum of stylistic confusion, and leaving editing and other upkeep a simpler matter.

The second set of guides was designed to cover any other miscellaneous subject, no matter how small or large, in whatever format seemed appropriate. These guides we dubbed "Miscellaneous Subjects Research Guides"; this series has no further relevance to present considerations, and all remaining discussion will refer to the DRG group.

The DRG are organized internally into six sections each. The first section lists some 10-30 or more subject readings pertaining to Library of Congress Subject Headings, adding their respective range of call

numbers coverage as an aid to shelf browsing in the reference and circulating collections. Section two lists a few to a dozen or more indexing, abstracting, and full-text services useful to finding articles in the periodical literature relevant to that particular department's subject interests; this section is broken down into listings for first-year students and non-majors, and for advanced researchers. Section three addresses, as relevant, leading hardcopy reference sources. The fourth section provides a link or links to that department's Web page, and its faculty descriptions. The fifth section sets out a hyperlinked list of professional societies and associations connected with that department's interests. The last section links to anywhere up to a hundred or more sites on the free Web that should be useful to the faculty and students in that department.

An important element of the DRG program is its joint administration by library personnel and the respective teaching departments involved. The first four sections of the guides are set up more or less entirely by the librarian liaisons to the departments, but the fifth and sixth are touted as "recommended by Department of \_\_\_\_\_ faculty," and in fact all materials that are added to them are at the very least passed by the departments for their approval before they appear online. This important element of the program is designed to help keep the latter more involved with the process, and the mere fact that it is set up that way is both good public relations for the library, and a true form of outreach.

### ***MONITORING THE PROGRAM***

To maintain stylistic consistency across the nearly 50 guides in the series, for the first several years, online maintenance of the html files was carried out by one office (the author and his assistant). This included all changes and additions to the series, and periodic checking of all the guides for broken links. The latter operation is carried out about once every two or three months to ensure currency. It is the responsibility of all the liaison librarians and their respective teaching departments to continue to provide new sites for inclusion in the lists.

The biggest challenge to making the program work has been getting all the librarians and other campus faculty to continue to supply materials for inclusion; obviously, the guides should optimally not only be current in terms of their connection to active sites, but also current with regard to continuing incorporation of the best materials relevant to that discipline. This has been an uphill battle at WKU, as library administration here has so far been unwilling to make any demands on all personnel toward this end.

In 2002 we decided it would be a good idea to start monitoring the number of “hits” on each guide. Therefore, I asked our campus computer services people to begin to keep counts of such accesses, and to report back to me with monthly totals at the end of each fiscal year. This they did, and my first report arrived in early June 2003. All of the guides had been accessed at least several hundred times, with the largest usage exceeding 1,500 accesses. The total figure of about 36,000 hits ended up exceeding the number of reference desk interactions over the same fiscal year period, so I began to think about how we might encourage even more involvement, and make use of the tallies of the numbers collected in additional contexts.

Encouraging more use at this point proved to be a daunting problem. Although it was possible to advertise the existence of the guides to faculty and staff through mass e-mails across campus, the university has no parallel mass function for contacting students—and these, of course, are the people we would expect to be the main users of the service. Counting on faculty to relay to their students the fact of their existence, and how they could be used to help their studies, is unfortunately in most instances not being very realistic.

It also became increasingly difficult to get the other librarians and library administration to further buy into the effort, since it was still viewed by many as the author’s—not Reference Services’ in general—“project.” It did help to suggest that the guides could be used as the starting point for research instruction classes, and a number of librarians have been using them in this connection. But this only accounts for the smallest percentage—undoubtedly under one percent—of all the classes being run at the university, and relatively very few students could be reached this way. Neither was it possible to convince those in control of the Library’s homepage on the Web to apply much additional space to promoting the service.

It was at this time that I realized that the statistics produced by number of Web page accesses might be used in a manner that could serve several important purposes at once, and proceeded to investigate.

### ***ANALYSIS, ASSESSMENT, AND PLANNING: THE PHILOSOPHICAL BASIS***

Around July 2003, after seeing the first fiscal year’s data on usage, I realized that our decision to make the DRG series department-based in the first instance rather than subjects-based made possible an important mar-

riage between service-oriented philosophy and statistical modeling. The usual approach to subjects-oriented research/resources guides addresses the matter of patron assistance in the same fashion that the old “pathfinders” concept did: By responding rather haphazardly to whatever individual subjects come to mind (as a function of the combination of probable patron interests, and librarian’s skills and knowledge). For this reason, coverage of subject matter is likely to be uneven, and worse yet, there is again no real way to assess whether all elements of the potential universe of library patronage are being equally served. It can be argued up front that the services of a university library, at least, should be allocated in such a fashion as to serve all elements of the campus equally (or at least equitably); further, one would like to have some means of actually measuring the degree to which this is true, and using such information to make additional efforts to reach those who are not being reached.

The DRG series concept is a perfect venue for such efforts because it is explicitly set up to address needs as measured on a department-by-department basis. The important linking concept here is that the number of accesses to the guides can serve as an important indicator as to how well the library is living up to its service responsibilities with respect to that interested population. Specifically, all other things being equal (i.e., with respect to the quality and orientation of the guides, and their relative degree of promotion), the number of visits to each guide should be directly proportional to the size of the potential user population.

This is an important point, because it lends itself to a form of statistical analysis that can be used to underwrite assessment efforts leading to service adjustments on an ongoing basis. If indeed it can be shown that particular guides are “underperforming” (i.e., they are receiving too few online visits), on the philosophical basis projected here there can only be two reasons: (1) they are not being promoted strongly enough, and (2) they are in some sense inherently inadequate with respect to amount or orientation of content. It turns out that this understanding easily can be translated into an inferential statistical model employing multiple-regression techniques, and the results from this, especially the residuals produced, can supply an unambiguous—and indefinitely repeating—form of assessment.

Using the data from the 2002-2003 fiscal year, I did a preliminary analysis that confirmed the feasibility of the approach. I decided to wait for another year of data, however, before describing the model in detail, as is now being done. In the meantime, the approach was integrated into the Library’s internal assessment plan.

The university here experienced its main decennial accreditation review in 2005, and as part of that process each department on campus extensively reviewed and revised its plans for internal assessment. In the case of the Library, such efforts are especially problematic because it is hard to come up with means of “closing the loop” in a planning sense that do not reduce merely to setting particular goals—often on an arbitrary basis only—and trying to achieve them. Instead, it is preferred to have stated, logically derived objectives, and statistical measures that can be brought to bear to meaningfully evaluate them.

The DRG program described here is eminently suited to providing such “closing the loop” planning because it can identify, indefinitely, those elements of the campus population that are being underserved, and suggest specific means for dealing with the problem. There are two distinct sides to the process.

First, and assuming that serious enough efforts are made to improve the service, the overall trends of success can be measured over a several-year period on the basis of the degree to which the numbers show an actual movement toward the “more potential users/more actual usages” standard. The multiple-regression approach to be described below can in fact produce easily interpretable results in this way: both with respect to the total number of accesses of the guides increasing, and the relative rates of usage across all the guides more closely approaching the ideal distribution. Actual movement in this direction should satisfy any external reviewer that the loop is in fact being closed in this approach, since not only could it be argued that we are better engaging our potential patronage in absolute terms, but that we are continually making efforts to adjust our services to ensure equity.

Second, the results of each year’s analysis can be used to plan outreach efforts to departments based on those results. This matter will be discussed later.

### ***ANALYSIS, ASSESSMENT, AND PLANNING: THE REGRESSION MODEL AND ITS USE***

Going through an actual analysis of the year totals in some detail will illuminate just how information of value can be obtained in connection with them.

Multiple regression is a multivariate statistical technique in large part based on correlation analysis. In simple correlation analysis, the variation in values of the cases making up two variables are related to one an-

other in a manner not calling for implied causality; that is, the association is posed in statistical probability terms alone. In simple linear regression, the model of possible correlation is extended to embrace the notion that one of the two variables, the dependent variable, represents some outcome that is to one degree or another caused by the operation of the other variable, the independent variable. In multiple-regression analysis, there is still only one dependent variable, and the values of its elements are projected to be at least in part caused by the combined influence of two or more independent variables.

Here, the dependent variable is the number of accesses of the guides over the fiscal year period. I have employed two sets of independent variables, one consisting of four (sub-)variables accounting for the varying sizes of the potential pools of users associated with each department on campus (for example, the number of undergraduate majors in each department), and a second consisting of three (sub-)variables accounting for variations in the size of the guides themselves (a coarse means of getting at the effectiveness of the guides: one supposes that a guide with less information provided will inherently deter usage—especially re-usage—as compared with one with more). The basic approach was to enter the independent variables into the analysis in two steps (this is, therefore, an example of what is known as “stepwise regression”), step one involving the four “size of potential user pool” variables, and step two the three “size/effectiveness of guide” variables. The residuals from the analysis were thus examined at two stages, after each step was completed.

In this instance (as is the case in many social science applications of regression analysis) the matter of focal interest was residual size. Remember that we are dealing here, most basically, with a correlation procedure; if the correlation is not a perfect one it means that the independent variables do not in combination fully explain the values of the elements of the dependent variable, and we are left with differences between their actual values and what the model estimates they should be. Sometimes the estimated value is higher and sometimes lower, the difference between the estimate and the actual being known as a residual—which, therefore, is sometimes a positive number and sometimes a negative number.

The basic results of the analysis after each of the two major “steps” are given as Table 1.

Here we see the original values of the elements of the dependent (“y”) variable, “accesses,” in the first column, the estimated value of y produced from the combined forces of the four “size of potential user



TABLE1. Multiple Regression Analysis Results

Departmental Guide	Y <sup>1</sup>	Y Estim. 1 <sup>2</sup>	First Residuals	Y Estim. 2 <sup>3</sup>	Final Residuals	Residual Change
Accounting	734	551.0	183.0	539.1	194.9	-11.9
Administration/Leadership	462	547.7	-85.7	516.7	-54.7	-31.0
Agriculture	540	587.6	-47.6	606.7	-66.7	19.1
Allied Health	487	448.0	39.0	472.1	14.9	24.1
Anthropology	403	425.7	-22.7	428.2	-25.2	2.5
Architecture/Manufacturing	464	514.2	-50.2	522.1	-58.1	7.9
Art	435	526.5	-91.5	496.7	-61.7	-29.8
Biology	726	612.6	113.4	683.9	42.1	71.3
Chemistry	528	488.8	39.2	518.1	9.9	29.3
Communication	561	561.2	-0.2	532.9	28.1	-28.3
Communication Disorders	488	512.5	-24.5	476.8	11.2	-35.7
Computer Science	655	553.8	101.2	561.0	94.0	6.3
Counseling	423	602.9	-179.9	554.8	-131.8	-48.1
Economics	495	502.1	-7.1	497.6	2.6	-4.5
Elementary Education	1076	970.6	105.4	948.1	127.9	-22.5
Engineering	519	520.4	-1.4	533.1	-14.1	12.7
English	717	607.8	109.2	644.5	72.5	36.7
Exceptional Children	833	556.4	276.6	524.8	308.2	-31.6
Finance	373	494.3	-121.3	464.8	-91.8	-29.5
Folk Studies	394	415.5	-21.5	420.1	-26.1	4.6
French	342	407.7	-65.7	379.8	-37.8	-27.9
Geography/Geology	590	507.6	82.4	605.8	-15.8	98.2
German	346	407.4	-61.4	379.1	-33.1	-28.3
History	596	624.9	-28.9	688.9	-92.9	64.0
Information Systems	357	507.0	-150.0	460.8	-103.8	-46.2
Journalism/Broadcasting	576	877.8	-301.8	925.5	-349.5	47.7
Library/Media Education	520	464.2	55.8	509.4	10.6	45.2
Literacy	347	420.4	73.4	392.7	-45.7	-27.7
Management	476	632.9	-156.9	606.4	-130.4	-26.5
Marketing	382	553.0	-171.0	512.1	-130.1	-40.9
Mathematics	634	505.4	128.6	526.3	107.7	20.9
Middle & Secondary	469	565.9	-96.9	533.5	-64.5	-32.4
Grades Military Science	588	422.8	165.2	430.1	157.9	7.3
Music	383	488.4	-105.4	477.1	-94.1	-11.3
Nursing	1197	700.2	496.8	713.0	484.0	12.8
Philosophy	648	420.2	227.8	445.4	202.6	25.2
Physical Ed & Recreation	492	585.5	-93.5	554.7	-62.7	-30.8
Physics/Astronomy	344	434.4	-90.4	449.7	-105.7	15.3
Political Science	441	531.2	-90.2	536.0	-95.0	4.8

TABLE 1 (continued)

Departmental Guide	Y <sup>1</sup>	Y Estim. 1 <sup>2</sup>	First Residuals	Y Estim. 2 <sup>3</sup>	Final Residuals	Residual Change
Psychology	475	666.4	-191.4	634.6	-159.6	-31.8
Public Health	502	546.7	-44.7	563.8	-61.8	17.1
Religion	456	455.3	0.7	457.7	-1.7	2.4
Social Work	566	481.0	85.0	496.9	69.1	15.9
Sociology	789	589.5	199.5	593.7	195.3	4.2
Spanish	404	437.8	-33.8	419.1	-15.1	-18.7

1. The dependent variable, total number of accesses of research guide, July 1, 2003 through June 30, 2004.

2. The combined estimate calculated from the four "user pool size" variables (number of undergraduate majors in department 2004, number of graduate students in department 2004, total student credit hours recorded in department's courses 2003-2004, number of full-time equivalent faculty in department 2003-2004; simple correlation coefficient  $r$  of each variable with the dependent variable: .555, .210, .276, and .298, respectively). At this point the regression model explained .368 of the total variation (multiple correlation of  $r = .606$ ) and was significant at  $\alpha = 0$ .

3. The final estimate, including the three "size/quality" variables (file size in kb, number of section two hyperlinks, number of sections five and six hyperlinks; simple correlation coefficient of each variable with the dependent variable: .281, .140, and .208, respectively). The final regression model explained .428 of the total variation (multiple correlation of  $r = .654$ ) and was significant at  $\alpha = 0$ .

pool" surrogates in the second column, and the residual value produced by subtracting the latter from the former in the third. Column four then relays the estimate of the remaining variation by the three "size of guide" variables, with column five relaying the final residual values produced by subtracting the latter from the former. The sixth and final column exposes one final piece of information: the change in residual values taking place between the end of execution of the first and second steps of the overall analysis.

This is not the place to dwell on various other statistical aspects of the analysis beyond stating that overall the seven variable model explained 42.76% of the variation in the data. This is not equivalent to a high correlation by natural science standards, but represents a perfectly respectable model for our purposes, since it is statistically significant with a 99%++ likelihood of representing an actual causal relationship, as opposed to one due to mere chance associations.

To understand how these numbers may be applied to real world considerations let us now look closer at Table 1.

The combined predictive effect of the first set of variables, measuring "size of potential user pool," is to produce an estimate of the actual values; again, the difference between the actual and the estimated values constitutes the first set of residuals displayed in the third numerical column in Table 1. A bit of examination shows that there are six residual

values in this column with values greater than 150.0, with the most extreme value, 496.8, being associated with the nursing guide. These guides were accessed a substantially greater number of times than the overall model would have predicted, and are thus “overachievers.” Conversely, there are also five residual values that are less than  $-150.0$ , including journalism/broadcasting, the lowest, at  $-301.8$ . These guides (and all the other ones leaving negative residuals) are relative “underachievers,” that is, based on the known sizes of the potential pools of users the overall model predicted they would have received more hits than they actually did.

These results are in themselves very informative, but they do not yet take into account the possible effect that varying size and quality of guides might have on their usage. Thus, the second set of (three) variables related to this matter are next added into the model, in “stepwise” fashion. Columns four and five show the new, adjusted, model estimates and resulting residuals—that is, the final residuals.

In the final residuals column we see that some of the extreme negative residuals have been reduced such that there are now only two cases of values more extreme than  $-150.0$ , but that there are still six instances of residual values that exceed  $+150.0$ . The explanation for these circumstances is as follows.

While it is the case that the addition of the last three explanatory variables to the model has in general added to its statistical explanation of the variation within the dependent variable (i.e., produced a higher multiple correlation between the dependent and independent variables), each individual case (research guide) will be affected uniquely. In the instance of the “underachieving” guides for counseling, management, marketing and psychology, it turns out that these score rather low on the three “size/quality of guide” measures. In theory, this helps to explain why these guides received fewer hits than did some others, in turn producing a combined estimate which is more on target than when one considers size of potential user population alone. The situation with the remaining underachieving guide, journalism/broadcasting, is different: not only does the size of the potential user population suggest it should be receiving more hits, but on top of this it scores above average on the size/quality measures. As a result, the final model ends up doing an even worse job of predicting its usage, with the final residual value being more extreme, instead of being reduced.

Meanwhile, it turns out that the addition of the new variables to the model has not much affected the level of prediction of the usage of the highly “overachieving” guides. Prediction would have been improved

had these scored high on the size/quality variables, but none did (in fact, most scored below average). Thus, the overachievement exhibited by these guides is likely due to some special act of promotion of their use, perhaps in connection with assignments made by that department's teaching faculty. Note, however, that not all of the guides that exhibited "overachievement" at the end of the application of the first round of independent variables were unaffected by the addition of the second round: for example, there was a dramatic change in the residuals left for the biology and geography/geology guides, both of which scored very high on the size/quality measures.

### *DISCUSSION*

It should be apparent by now that it is the purpose of the multiple-regression analysis to expose those guides that have been, relatively, underachievers or overachievers in attracting use by the potential user population. At this point, the emphasis shifts from mere measurement to a consideration of what to do with this information in a practical sense. Herein lie opportunities for any number of creative actions designed to bring the information on the guides to their intended audience. If, for example, the data suggest a particular guide is underperforming for reasons of lack of promotion, then means for improving such could be investigated. One could distribute generic e-mails to faculty or students extolling the usefulness of the guide, or suggest to faculty that guides form the base component of research instruction sessions run by either, or both, that faculty member or a librarian. Or one could extract suggestions for improvement through focus-group sessions with groups of students, or individual consultations with faculty members. At the same time, an examination into the "overperforming" guides might expose some of the reasons for their better performance: for example, that more faculty members in such departments are assigning term papers or integrating a library research component into their classes.

Conversely, if there are significant positive residual shifts between the first and second sets of results within the stepwise regression, the quality and/or emphasis of the guide needs to be re-examined: for example, perhaps it doesn't contain enough information to make it a useful tool (e.g., number of links to free Web sites in that subject), or perhaps it features the wrong kind of information. Again, interviews with faculty or focus group sessions might provide answers in this regard. It might well turn out, for example, that students would prefer to have more jobs-related and

technically related tutorial sites included within the sixth section of the guides, or that faculty members think the same section should expand its coverage of news-related sites concerning that subject area.

Before closing, a couple of possible objections to this means of meta-analysis should be refuted. First, it is sometimes said, with reason, that access tallies alone don't tell us why individuals are visiting a particular Web site, or whether they are satisfied with what they find. In the present instance, however, this is a weak criticism, because it is a reasonable assumption that on the average the number or percentage that access but do not make effective use of each guide will be at least roughly the same from guide to guide. Otherwise put, it is not the absolute number of hits that matters here, just the relative numbers. Ideally, it would be nice if everyone became fully engaged in each guide's content each time they visited, but the analysis model is still applicable even if they do not.

It might also be objected, independently, that even though the analysis model appears valid in a technical sense, we are still left with no overall assessment of the guides' adequacy. To this I respond that in fact we are, as we are provided each year with two important statistical indicators of such adequacy: (1) the absolute numbers of hits themselves (which should at the very least continue to exceed some nontrivial rate of usage, without declining), and (2) the amount of variation explained by the regression model (which, if we are on the right trajectory over time, should both increase in absolute terms, and become more and more tied to the size of potential user pool variables than to the "quality" variables).

One can imagine, therefore, a continuing assessment cycle in which "underperforming" guides are identified, the causes of such underperformance investigated and dealt with through various adjustments, and the adjustments themselves investigated anew through study of the following year's data. In this way the guides become a dynamic form of evaluation of the online reference service—one which accumulates data as time goes on, and periodically exposes relative weaknesses due to dips in quality or publicity efforts.

## **CONCLUSION**

It should be emphasized that the analysis model described here makes sense only when applied within a context of measurable surrogates for size of potential usership. The latter are hard to develop if not directly a function of discrete entities such as academic departments. Still, as long as one keeps within this basic limitation it is difficult to imagine a "clos-

ing the loop” kind of assessment model and procedure that is more elegantly to the point than this one is. Additionally, once one has collected the basic data, the analysis itself is not terribly time consuming: I performed the entire regression analysis by hand on a pocket calculator in less than one day. [I might add that librarians should not feel intimidated by the intricacies of inferential statistical techniques a priori, especially if they wish to have the work they are pursuing truly considered library science. Almost all statistical packages these days—SPSS (Morgan 2004) and others—provide convenient means for doing multiple regression and other operations, and neophytes might want to take a look at Vaughan (2001) or other texts (e.g., Anderson 2003; Tabachnick and Fidell 2000; Walsh and Ollenburger 2001) to help them appreciate such techniques. Further, and although few library schools offer courses in inferential statistics, those librarian colleagues who have additional graduate training in other fields can often be tapped for assistance—not to mention the institution-wide services that almost all colleges and universities establish specifically to assist their personnel in the methodological aspects of research.] Further, there is the advantage that only the most underachieving guides need be given any special amount of attention in any given year, and even this need not involve a total of more than several hours of labor per guide per year.

There is, however, at least one potential Achilles heel in this kind of project that needs to be addressed at its outset. This is the matter of who to answer to. Experience here suggests that the various librarians charged with collecting information for the guides are not likely to become adequately engaged unless the entire effort is run under the direct supervision of the library’s senior administrative official (or at the least, its person in charge of all user services). This person will need to become proactive in ensuring that changes/improvements in the guides are actually made when called for.

In the passage from Peters (2000) reproduced earlier that writer anticipated a form of meta-assessment contributing to an “automatically customized or self-improving” information environment. The system reported here comes very close to approaching this ideal.

## REFERENCES

- Anderson, Theodore W. *An Introduction to Multivariate Statistical Analysis*, 3rd ed. (Hoboken NJ: Wiley-Interscience, 2003).
- Butfenfield, Barbara. “Usability Evaluation of Digital Libraries.” *Science & Technology Libraries* 17 (1999): 39-59.

- Carson, Bryan, and Charles H. Smith. "Integrating the Library into Your Classroom." Building a Scholarly Structure for Teaching and Learning (conference), Bowling Green, KY, 22 May 2001.
- Morgan, George A. *SPSS for Introductory Statistics: Use and Interpretation*, 2nd ed. (Mahwah NJ: Lawrence Erlbaum, 2004).
- Nichols, James O., and Karen W. Nichols. *The Departmental Guide and Record Book for Student Outcomes, Assessment, and Institutional Effectiveness*, 3rd ed. (New York: Agathon Press, 2000).
- Peters, Thomas A. "Current Opportunities for the Effective Meta-Assessment of On-line Reference Services." *Library Trends* 49 (Fall 2000): 334-349, on p. 347.
- Smith, Charles H., Bryan Carson, and Haiwang Yuan. "A Reference Outreach Project Integrating Departmental Faculty Input." KLA Annual Conference, Louisville, KY, 18 Oct. 2000.
- Tabachnick, Barbara G., and Linda S. Fidell. *Using Multivariate Statistics*, 4th ed. (Boston: Allyn and Bacon, 2000).
- Vaughan, Liwen. *Statistical Methods for the Information Professional: A Practical, Painless Approach to Understanding, Using, and Interpreting Statistics* (Medford, NJ: Published for the American Society for Information Science and Technology by Information Today, 2001).
- Walsh, Anthony, and Jane C. Ollenburger. *Essential Statistics for the Social and Behavioral Sciences: A Conceptual Approach* (Upper Saddle River, NJ: Prentice Hall, 2001).

doi:10.1300/J120v47n97\_08