

UC Merced

UC Merced Previously Published Works

Title

Evaluating Library Instruction: Doing the Best You Can With What You Have

Permalink

<https://escholarship.org/uc/item/7z41p7k8>

Journal

RQ, 33(2)

Author

Barclay, Donald A.

Publication Date

1993



Library Literacy

Mary Reichel, Editor

In this column, Donald Barclay discusses "evaluation in the real world" and describes a successful library instruction evaluation effort he conducted for freshman writing students. His basic points are forceful: instruction librarians need to evaluate students' learning in library instruction programs, and evaluation efforts need to be realistic in terms of time and resource limitations.

Barclay's approach is worthwhile. He

gives a model for evaluating library instruction which librarians on the front lines can use and modify. His test is simple, but it addresses the major learning goals most instruction librarians have when working with beginning students. He also gives inspiration for librarians to begin evaluation efforts and data gathering even if the projects are not comprehensive research studies. If evaluation provides useful local data, that may be a sufficient result.—*Ed.*

EVALUATING LIBRARY INSTRUCTION: DOING THE BEST YOU CAN WITH WHAT YOU HAVE

DONALD BARCLAY

Donald Barclay is Coordinator of Instruction at the New Mexico State University Library, Las Cruces.

"Had we but world enough, and time. . . ."—Andrew Marvel

Calls for Evaluation of Library Instruction

Just as Beals, writing in 1942, pointed the finger at all librarians for the lack of hard research in the library field, Werking, writing in 1980, took teaching librarians to task for the general lack of meaningful evaluation of library instruction programs.¹ Since 1980, a number of articles expressing simi-

lar criticisms of library instruction have appeared in the literature. In 1982, Hardesty, Lovich, and Mannon wrote of library instruction that "there is a good deal of talking about evaluation, but few seem to be doing anything about it."² In that same year, Cottam wrote that when there is evaluation of library instruction, it is usually conducted as an "afterthought" instead of being planned into an instruction program.³ More recently, library-instruction bogeyman Tom Eadie complained that most evaluations of library instruction focused on user satisfaction rather than on what students really learned.⁴ And, like it or not, the statistics support Eadie. A 1989 survey by Mensching found that only 62 percent of responding library instruction programs did any kind of evaluation at all, and only 23 percent of the responding programs used testing for evaluation.⁵ Even worse, a more recent survey found that evaluation of li-

Mary Reichel is University Librarian at Appalachian State University, Boone, North Carolina. Correspondence concerning this column should be addressed to Mary Reichel, University Librarian, Belk Library, Appalachian State University, Boone, NC 28608.

brary instruction was actually less common than Mensching's study suggests.⁶

Evaluation in the Real World

Why are teaching librarians so unwilling or unable to conduct evaluation? Is it because, as the Tom Eadies of the library world suspect, that teaching librarians are selling snake oil and thus fear an impartial evaluation? I think not. As Sugrañes and Neal have pointed out, the simple, un sinister reasons teaching librarians tend to avoid evaluation are that evaluation is seen as too complex and too time consuming.⁷ This is especially true for those teaching librarians (94.6 percent according to a BIS survey) whose duties include more than library instruction.⁸ With so much to do, it is not surprising that when the work-a-day (and occasional nights and weekends) teaching librarian must choose between getting more tasks accomplished—more classes taught, more hours served on the desk, more books selected for the collection—and evaluating a task that has already been completed, evaluation comes last. As with most library problems, the lack of evaluation of library instruction is a matter of too few resources being stretched too far.

Of course, just because teaching librarians are busy people doesn't mean that demands for evaluation of library instruction will go away. With the recent emphasis on outcomes assessment in higher education, directors, deans, university presidents, and even state legislatures are increasingly wanting to know just what college students are learning. Because of this trend, the existence of many library-instruction programs may well rest on the ability of teaching librarians to evaluate instruction and come up with meaningful data that show just what students have learned. Increased demands for evaluation, however, are not likely to be met with increased support for conducting evaluation, leaving teaching librarians in a Catch-22 where increased support for instruction depends on evaluation, but evaluation can't be conducted without increased support.

And consider how much support it takes to do an evaluation equal to, let's say, some gem of a library-instruction evaluation published in a big-time library journal. The one authored by the full-time head

of a huge instruction program at an ARL library. The one bursting with chi-squares and tables. The one with an *N* equal to more *S*s than most teaching librarians will see in two years. How can the average teaching librarian match that? When will the average teaching librarian get the kind of support necessary to carry off such a complex evaluation? The answers to these questions are simple: "You can't!" and "Never." But these are answers to the wrong questions. The real question should be "How can the average teaching librarian conduct meaningful evaluation?" The answer to this is also simple: "Set your sights lower and do the best evaluation you can with what you have."

This answer is not based on the premise that teaching librarians should do sloppy research; on the contrary, it demands the best research possible. This answer is, however, based on the premise that some hard evaluation data, even if the data may be less than perfect, are better than either no data at all or soft data obtained from anecdotal observation and surveys of student satisfaction.

What follows is a discussion of methods of evaluation and their practicality for real-world teaching librarians. Along with this discussion is a description of an evaluation I conducted. This evaluation—admittedly less than perfect and subject to improvement—was one that I was able to conduct with the resources available to me, a somewhat average instruction/reference librarian. Some will be able to conduct much better evaluations; others will never have the resources to do even what I have done. Fine. My purpose here is not to hold up this evaluation as an ideal; instead, I offer it as an example that will, I hope, encourage other teaching librarians to conduct some kind of meaningful evaluation by using whatever resources they have at hand.

How to Evaluate?

Typically, there are four methods for collecting information to evaluate the effectiveness of instruction: anecdote, survey, test, and evidence of use. Collecting information through anecdotal observation is something librarians do every day. While anecdotal information is important—after all, it is a lifetime of accumulated anecdotal information that makes

the experienced librarian valuable—anecdote is subject to personal bias and is often considered soft and unreliable by outsiders (directors, deans, accreditors, etc.). Survey information—most typically obtained by asking students how helpful instruction has been to them—is useful for judging student perceptions, but it has shortcomings in that students, especially basic library users, may often overrate or underrate the value of what they have learned. Like anecdotal data, satisfaction survey data are not highly valued by outsiders. Though often collected by instruction librarians, neither anecdotal information nor survey information by itself produces hard evaluative data. Testing, on the other hand, has the value of determining what students have learned as opposed to how they feel about what they have learned. The validity of test-based data is, however, entirely dependent on the quality of the test itself. If success on a test does not equal successful library use, then the test data are invalid. This pitfall can be avoided when evidence of use is the basis for evaluation, but this method of evaluation requires either a great deal of cooperation from the subjects (who might be asked to keep research logs and submit final bibliographies) or sufficient resources to scientifically observe a significant number of subjects. For this reason, evidence of use is often not an option for the average instruction librarian.⁹

Of course the four methods of collecting data are not mutually exclusive. Testing and surveying are often done at the same time and might be followed up by examining bibliographies (evidence of use). Also, any of the above methods can be enhanced by comparison between the success of those who received instruction and those who did not (the control group). Control groups are often a problem in library instruction, since the only way such a group can be created is to deny instruction to certain students (which is unethical) or to enlist the cooperation of those who choose not to take advantage of library instruction (which may be impossible). As Frick has pointed out, most librarians have “little opportunity to set up control groups (even ignoring the troublesome question of the ethics of doing so), and virtually no opportunity to select students based on their

backgrounds (to control for the level of acceptable experience).”¹⁰ Faced with the choice of evaluating without a control group or doing no evaluation at all, instruction librarians should not be afraid to proceed without a control group, even though this is not ideal.

Test-based Evaluation

For the reasons discussed above, testing is often the only practical way for instruction librarians to collect hard evaluation data. When choosing what kind of test to use, instruction librarians do not have to choose between a standardized test and a locally produced test. The reason for this is that there are no widely accepted standardized tests for evaluating library use on the college level. This is unfortunate because standardized tests, despite their drawbacks, are easy to administer and often carry more credibility with outsiders than do locally produced tests.¹¹ Given the lack of appropriate standardized tests, the instruction librarian who wishes to evaluate by testing has no choice but to create a local test or to adopt someone else’s local test.

When creating a test the instruction librarian will usually have to choose between a multiple-choice or free-response format. Multiple-choice has the advantage of being easy to score and would be a wise choice for many librarians working with limited resources. However, multiple-choice has been criticized as being “limited to well-structured problems,”¹² and library research is certainly not a well-structured problem. The act of writing an answer to a free-response question, on the other hand, has more in common with the unstructured act of library research and so may be a better test of a student’s ability to use a library. The drawbacks of free-response tests are that they are difficult to score and are often seen as less objective than multiple-choice tests.

The debate between the validity of “objective” multiple-choice testing versus the validity of “subjective” free-response testing is too big to be covered here. It is worth noting, however, that Frary’s simulation study of multiple-choice versus free-response concluded that “both reliability and validity of free-response scores are moderately higher than those of multiple-choice scores.”¹³ Frary also noted that

studying the value of multiple-choice versus free-response is a problem because specific research on this subject is limited.¹⁴ In addition, the reliability and validity of free-response testing depend a great deal on how the tests are scored. One simple technique to improve the validity of free-response tests is to base scoring on a list of the key concepts that should be expressed in an ideal written response. Whenever a student expresses a key concept in a written response, a fixed number of points is then awarded to that answer. Another technique to improve free-response test validity is to use double-blind scoring in which two scorers working from the same list of key concepts score each test independently. The two scores are added together to produce the final score. If the two scores vary by more than an established amount, a third scorer can be called in. Of course such scoring is very labor intensive, so the librarian with limited resources may be wise to stick to a multiple-choice format. A third option—creating a test based on short-answer responses—can serve as a middle ground between multiple-choice and free-response, though short-answer tests have more in common with multiple-choice than with free-response tests.

How I Done OK

Weighing all the options, I chose to evaluate by testing combined with survey: testing, because I wanted to evaluate students' ability to use the library; survey, because student opinions of library instruction are helpful. Besides, it is a simple matter to attach a satisfaction survey to a test. Because a control group could not be created, I did my evaluation without one. Furthermore, I evaluated only freshman writing students (both honors and general-option). Freshman writing supplies the largest number of students to my library's instruction program and is the group most available for evaluation. Ideally, I would have surveyed randomly chosen students from all the different courses that receive library instruction, but that wasn't possible with the resources I had.

Once I decided on testing, I chose to use free-response questions despite the extra work this involved. My main reason for choosing free-response is my conviction that multiple-choice tests are too far re-

moved from the act of using a library. A second reason for using free-response was that the subjects of the evaluation were enrolled in a writing course and so free-response seemed more in keeping with the nature of the course. To construct the test I asked myself a simple question: "What are the main points I am trying to teach freshman writing students?" From my question came a brutally simple test (see appendix A). I could have spent a lot of time creating a more complicated test, but it still would have had flaws and certainly would have been harder to administer and score than the simpler test I settled on.

With the cooperation of the freshman-writing instructors, the test was administered as a pretest early in the semester, before freshman writing students had been brought in for library instruction or given a library assignment. Essentially the same test (along with some satisfaction-survey questions) was then administered as a post-test late in the semester, after students had attended a library-instruction session and worked on their major research papers (see appendix B). Using a list of key concepts and 0-5-point scale, I scored the pre- and posttest myself. A score of 0 means the answer was completely incorrect; a score of 5 indicates a completely correct answer.

I acknowledge the bias that had to result from scoring the tests myself. Certainly, my methods are a far cry from the ideal which calls for evaluation measures to be constructed, administered, and scored by someone independent of the library-instruction program.¹⁵ In my defense, I can say that had I waited for the luxury of an independent agent to evaluate my library-instruction program, I would still be waiting.

Once I had scored the tests and coded the numerical results on computer forms, the raw data were ready to process. I am fortunate that my institution has a Department of Experimental Statistics which is charged with helping faculty conduct statistical research. Because of this, it was no problem to have the raw data converted to useful statistical information. For other librarians, converting data to statistics may be the most terrifying research problem of all and, I suspect, may have as much as anything to do with causing library instructors to avoid evaluation. Solutions to

this problem include learning to do statistics yourself, finding a kind soul to do the statistics for you, or arranging with your library administration to provide some kind of support for this part of your evaluation. The last may not be hard to do, as it is collecting the raw data, not running it through a statistics computer program, that is the truly time-consuming part of research. If all else fails, anyone can figure simple percentages from raw data and infer from the percentages what they can. Figures 1 and 2 show the pre- and posttest scores as percentages and provide an adequate picture of the results of the evaluation.

What the Evaluation Told Me

First of all, the pretest confirmed something I already knew anecdotally—most freshmen know little about locating books by subject and practically nothing about locating periodical articles by subject. Second, the statistics generated by the comparison of the pretest with the posttest told me that the library and the writing program are doing a good job of teaching freshman writing students how to find books and periodical articles by subject. Comparison of the pretest and posttest for locating books by subject resulted in $t = 13.107$ and $p = 0.0001$; comparison of the pretest and posttest for locating periodical articles by subject resulted in $t = 12.21199$ and $p = 0.0001$. Are these good numbers the result of a less-than-perfect evaluation? Perhaps. I expect that as I refine and readminister this evaluation, the numbers may not turn out as good. But even if the current numbers are not perfect, they still indicate that students are learning the two essential information-finding techniques we want them to know by the time they finish freshman writing. And the numbers indicate this more accurately and with more authority than any amount of anecdote-based guesswork or soft data from satisfaction surveys.

What Has This Evaluation Done for the Library's Instruction Program?

Although this article has mentioned outcomes assessment and program justification as reasons for evaluation, the fact is that no evaluation should be undertaken

with only these goals in mind. Unless evaluation will somehow improve the thing being evaluated, it is not worth doing. The evaluation I conducted, despite its flaws, has helped the library-instruction program in several ways. First of all, the act of answering the evaluation questions on the pretest helped students by preparing them for what they would learn when they came to the library. Because the answers to the pretest questions were read by the writing instructors before passing them on to me, the writing instructors got a realistic picture of how much—or, more exactly, how little—their students knew about using libraries. The pretest answers were also read by instruction librarians prior to meeting with each freshman writing class, thus allowing the librarians to adjust instruction sessions to the students' level. For example, the pretests might warn the instruction librarian that half the students in an incoming class thought the card catalog was the best place to find a book by subject when, in fact, some 85 percent of the library's books (including all post-1970 books), are in our online catalog. As for the posttest answers, reading these helped both writing instructors and teaching librarians get a sense of what concepts did and did not stick with students, resulting in changes in what is emphasized in both library instruction sessions and library assignments. Further, the act of answering the posttest questions helped students by requiring them to recollect and express the library skills and strategies they had learned during the semester, thus helping to fix these skills and strategies in their minds. Even if it had never been turned into statistical information, the evaluation would have been worth the effort because it helped to improve library instruction and foster learning.

As statistical information the evaluation was useful in showing that freshman writing students had indeed learned the basics of using an academic library. The results of the evaluation were presented to library faculty, English faculty, and the Faculty Senate Library Committee, catching the attention of outsiders even though I was careful to point out the survey's flaws as I saw them. Again my point: less-than-perfect evaluation is better than no evaluation; hard data are better than soft data.

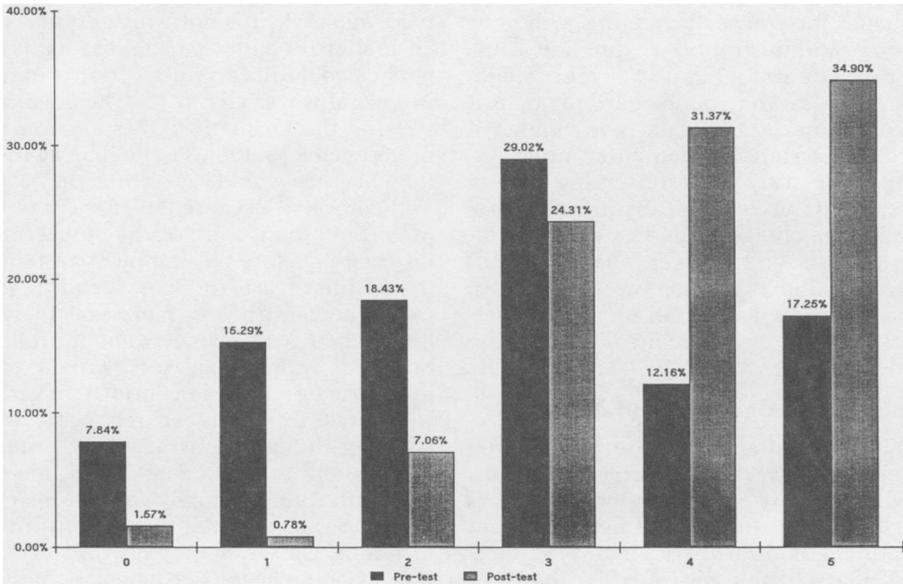


Figure 1. Finding Books by Subject

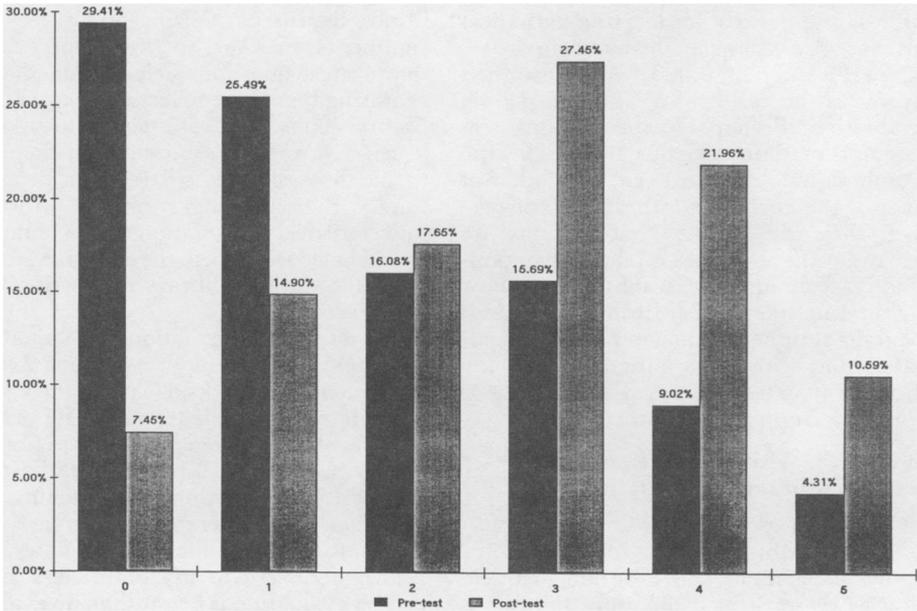


Figure 2. Finding Periodical Articles

How I'll Do Better

Having done test-based evaluation once, I plan to do it again, rectifying as much as possible the mistakes and short-cuts that made the original evaluation less than ideal. The biggest mistake I made the first time around was failing to ask on the posttest if the student had actually attended a library-instruction session; next time, this question will be asked so that the gains of those who came to the library for instruction can be compared to those who did not attend. Further, in the future I plan (1) to use double-blind scoring to cut down on scoring bias, (2) to separate the responses of the honors students from the general-option students, and (3) to study the correlation between satisfaction as indicated on the survey and learning gains as indicated by the evaluation. I am also considering using test-based evaluation on courses other than freshman writing. In short, I

plan to push my less-than-perfect evaluation as close to perfection as I can given the resources at my disposal.

Conclusion

Criticism of instruction librarians for their general failure to conduct meaningful evaluation, while valid, has failed to recognize the limited resources most teaching librarians have for conducting evaluation. In response to such criticism, instruction librarians need to conduct the best, most meaningful evaluations they can with the resources they have. Teaching librarians must then use the data from these evaluations to improve their instruction programs and to convince outsiders of the value of library instruction. That these evaluations may not meet the highest standards of scientific rigor should not be a deterrent, but instead serve as a spur to continually improving the quality of evaluation. ■■

REFERENCES AND NOTES

1. Ralph Beals, "Implications of Communications Research for the Public Library," in *Print, Radio, and Film in a Democracy; Ten Papers on the Administration of Mass Communications in the Public Interest, Read Before the Sixth Annual Institute of the Graduate Library School, the University of Chicago, August 4-9, 1941* (Chicago: Univ. of Chicago Pr., 1942), 180; Richard Hume Werking, "Evaluating Bibliographic Education: A Review and Critique," *Library Trends* 29 (Summer 1980): 153-64.
2. Larry Hardesty, Nicholas P. Lovich, Jr., and James Mannon, "Library-Use Instruction: Assessment of the Long-Term Effects," *College & Research Libraries* 43 (Jan. 1982): 38.
3. Keith M. Cottam, "Avoiding Failure: Planning User Education," *RQ* 21 (Summer 1982): 333.
4. Tom Eadie, "Beyond Immodesty: Questioning the Benefits of BI," *Research Strategies* 10 (Summer 1992): 108.
5. Teresa M. Mensching, "Trends in Bibliographic Instruction in the 1980s: A Comparison of Data from Two Surveys," *Research Strategies* 7 (Winter 1989): 11.
6. Otis Chadley and Jacquelyn Gavryck, "Bibliographic Instruction Trends in Research Libraries," *Research Strategies* 7 (Summer 1989): 111.
7. Maria R. Sugañes and James A. Neal, "Evaluation of a Self-Paced Bibliographic Instruction Course," *College & Research Libraries* 44 (Nov. 1983): 444.
8. Charles D. Patterson and Donna W. Howell, "Library User Education: Assessing the Attitudes of Those Who Teach," *RQ* 29 (Summer 1990): 519.
9. One type of observation-like evaluation that may be an alternative for the average instruction librarian is the focus group. For an example of the use of a focus group in a research library, see Richard Widdows and others, "The Focus Group Interview: A Method for Assessing User's Evaluation of Library Service," *College & Research Libraries* 52 (Spring 1991): 377-85.
10. Elizabeth Frick, "Qualitative Evaluation of User Education Programs: The Best Choice?" *Research Strategies* 8 (Winter 1990): 9.
11. Marsha V. Krotseng and Gary R. Pike, "Cognitive Assessment Instruments: Availability

- and Utilization,” in *A Practitioner’s Handbook for Institutional Effectiveness and Student Outcomes Assessment Implementation*, ed. James O. Nichols (New York: Agathon Pr., 1991), 87.
12. Robert L. Linn, “Achievement Testing,” in *Encyclopedia of Educational Research*, Vol. 1, 6th ed., ed. Marvin C. Atkin (New York: Macmillan, 1992), 7.
 13. Robert B. Frary, “Multiple-Choice Versus Free-Response: A Simulation Study,” *Journal of Educational Measurement* 22 (Spring 1985): 30.
 14. *Ibid.*, p. 21.
 15. Larry Hardesty, *Evaluation of Bibliographic Instruction* (Urbana, Ill.: ERIC Clearinghouse, ED 271 115, 1986), 9.
-

APPENDIX A
E-111/E-111-H LIBRARY INSTRUCTION EVALUATION

Student number: _____

1. Student X needs to find a book on the subject of _____.
List all the steps Student X should take in order to find and obtain such book in the NMSU Library.
 2. Student Y needs to find a periodical article on the subject of _____.
List all the steps Student Y should take in order to find and obtain such an article in the NMSU Library.
-

APPENDIX B
LIBRARY USE QUIZ

Student number: _____

1. You need to find a book on the subject of _____.
List all the steps you should take to locate and obtain such book in the NMSU Library.
2. You need to find a periodical article on the subject of _____.
List all the steps you should take to locate and obtain such an article in the NMSU Library.
3. What would you change about the library instruction session you attended as part of your English-111 class?
4. What did you like best about the library instruction session you attended as part of your English-111 class?

(Appendix B. Satisfaction survey questions. Post-test only.)