

Evaluating Bibliographic Education: A Review and Critique

RICHARD HUME WERKING

AS THE POPULARITY OF BIBLIOGRAPHIC instruction* has grown, concern about its evaluation has more than kept pace. Champions of instruction, as well as critics and neutral observers, have urged practitioners to evaluate their programs, or researchers to develop better methods of evaluation for others to use. Such concern has met with a response that in recent years has become noteworthy.

The appearance of a considerable number of books, articles and other documents over the years has failed to still the clamor. Although recognizing the recent increase in the number of such pieces, Brewer and Hills observed in their 1976 state-of-the-art review, "It is significant that there are few references to evaluation in the literature of reader instruction and until very recently they have been virtually non-existent."¹ Likewise, Fjällbrant in 1977 approvingly quoted Lubans: "Instructional programs in all types of libraries have been infrequently evaluated; their need and effect have not been measured except in a few isolated cases."² A critic of bibliographic education has delivered the same message, in language quite unlike what is generally heard at conferences of instruction librarians or found in the library literature.

Richard Hume Werking is Assistant Director, Reference and Collection Development Services, University Libraries, and Assistant Professor of History, University of Mississippi.

*For purposes of stylistic relief, the terms *bibliographic instruction*, *instruction*, *bibliographic education*, and *user education* appear interchangeably in this essay, although I prefer the latter two terms.

According to Benson: "Bibliographic instruction seems to be perceived by many librarians simply as a self-evident social good, not needing an extensive rationale or empirical evidence to substantiate its effectiveness or even to support the need for it. Much of the literature of bibliographic instruction resembles a dialectic with the antithesis missing."³

General complaints about the lack of evaluation at the program level have been substantiated with specific evidence. In a survey reported in 1975, Peter Hernon found that about two-thirds of responding libraries were not collecting data with which to review their library lecture programs.⁴ Likewise, James Ward's survey of instruction programs in southeastern academic libraries revealed that more than three-quarters were not using any evaluative instrument.⁵ Even some of the better-publicized programs of bibliographic instruction, supported with grants from the National Endowment for the Humanities and the Council on Library Resources, have evidently done little to evaluate their efforts.⁶

Such is the theme that emerges from the literature. It is a message that will likely persist for some time, unconquered if not undaunted by the appearance of articles outlining the evaluation of user education programs and techniques at such places as the University of Arizona, Brigham Young, DePauw, Pennsylvania State, or the Wooster Agricultural Technical Institute. Before elaborating on this theme, this survey of the literature will examine reasons for evaluating, what and how instruction librarians evaluate, problems with evaluation, and questions of proof.

Why, What and How to Evaluate

There are few explicit disagreements about the definition of formal "evaluation." Suchman has characterized it as "an appraisal of value," while others have stressed its role in describing outcomes as well as placing values on them.⁷ Fjällbrant provided a succinct description covering both emphases; she observed that the "purpose of evaluation is to collect and analyze information that can be used for rational educational decision-making."⁸

Systematic evaluation in user education occurs for a number of reasons. Surveying 136 instruction librarians at liberal arts colleges, Lindgren found that 90 percent of the 68 who evaluated did so to improve the instruction program. The next most popular reason, cited by 43 percent, was "to justify the program to oneself." Other reasons found by Lindgren all involved justification of the program to various

Evaluating Bibliographic Education

elements: the faculty (40 percent of the evaluators so indicated), the college administration (31 percent), and the library administration (18 percent).⁹ Lindgren's list by no means exhausts the possibilities; additional reasons noted by other investigators include gathering information as part of a needs assessment, comparing different instructional methods, defining and redefining goals and objectives, reinforcing students' learning, and gaining "visibility" for the library.¹⁰

The objects of evaluation also vary. Instruction librarians agree in general that their efforts are intended to provide students with the ability to use the library more efficiently and effectively than they would without instruction. But there is much less agreement about more specific instructional goals and also about the form instruction should take. There is similar, and closely related, disagreement about how librarians should systematically determine if either the general goal or more specific goals are being met. Consequently, there are several answers to the questions of what and how to evaluate.

Kirk has provided a useful taxonomy describing several ways of measuring student achievement in library use. One common method is to gauge the mastery of the *content* of bibliographic instruction by administering a test which asks questions about such items as the parts of a catalog card or citations from a journal index. Librarians can also examine the *product* of a student's endeavors in the library, such as the quality of a term paper's bibliography, and even the *process* of the student's library work—i.e., was it an efficient method?¹¹

Although improvement in library use skills (variously defined) is the most common object of bibliographic education and thus of evaluation, it is not the only one. Kirk, Fjällbrant, Vogel, and others stress the importance of changing attitudes as well. According to Kirk, instruction programs "must change the attitudes of library users into positive relationships or positive feelings towards the library and librarians," presumably because those positive relationships or feelings will contribute to the library's goals.¹² Thus, librarians frequently survey students about their feelings toward libraries and librarians, although reports of such surveys have constituted a relatively minor part of the evaluation literature.¹³ Other candidates for assessment are changes in patterns of library use (perhaps as indicated by circulation and reference statistics or the number of students doing research) and the position the instruction program holds among the library's and the institution's priorities.¹⁴

Virtually inseparable from what is evaluated is the question of how to evaluate systematically, going beyond the librarian's observation of

patron behavior, which is an integral and extremely important part of library service. Indeed, observation is no doubt the most common method of evaluation by librarians, and the basis for all sorts of opinions. However, it is not systematic.¹⁵ As noted above, librarians try to gauge learning by administering various tests, a few of them more or less standardized and others homemade. Of Lindgren's respondents, 43 percent used some sort of test, but his category "testing user performance" leaves much unclear about what was actually tested and how. More frequent, Lindgren found, was a survey of student attitudes, conducted by 74 percent of responding librarians. Here, too, it is not clear whether the survey was by questionnaire or by interview, although almost certainly the former method was heavily predominant. Less popular than the student survey was a faculty survey, undertaken by 60 percent of the respondents. One-quarter of his respondents measured performance in other ways.¹⁶

Those other methods of measurement focus chiefly on less obtrusive measures than tests, questionnaires or interviews. They include examining the products of instruction, such as term paper references and bibliographies.¹⁷ Also, to study the process of library use and the impact of instruction upon it, librarians and faculty occasionally ask groups of students to keep logs which track their library use in terms of such factors as sources consulted and in what sequence, and time spent on various activities and sources.¹⁸ Library use is also measured at a more "macro" level than the individual student, for example, by number of books circulated, number (and sometimes kinds) of reference questions asked, number of individuals entering the library, and number of interlibrary loan requests.¹⁹

If observers outside the circle of instruction librarians find these methods less than completely satisfactory, they are not alone. Kirk spoke for many of his colleagues when he declared, "We are all thoroughly dissatisfied with the kinds of evaluation tools available."²⁰ The method which is most criticized, and which is paradoxically that most commonly reported in the literature, is the written test. As already noted, there is a discrepancy between its relative infrequency in practice and its role as the dominant type reported in the literature.²¹ In all likelihood, practicing librarians feel more comfortable in constructing their own survey instruments to determine student and faculty attitudes than they do in constructing their own objective tests. Thus, there has been a much larger market for articles describing the development and use of tests.

The Test

The single most popular library skills test is the Feagley Test, developed in 1955 by Ethel Feagley and her associates at Columbia University. Several pages and eighty questions cover: parts of a book; definitions of terms such as "format," "italic" and "imprint"; arrangement of headings in the card catalog; and six other categories.²² The test was constructed as a diagnostic device to determine the level of library skills, rather than to assess the impact of bibliographic instruction.²³ Many schools, including Earlham, Lawrence, and Towson State as early as 1939, have used some form of written test to determine the most basic library skills level of incoming students, and occasionally to serve as a teaching device.²⁴ Such tests are usually much shorter and simpler than Feagley's.

The past few years have witnessed a number of reports in the literature detailing the construction and use of tests to measure the impact of bibliographic instruction. One large group consists of objective tests designed to determine, according to Kirk's taxonomy, the content of instruction. Parlett and Hamilton refer to this model as employing the "agricultural-botany paradigm" and summarize it as follows: "Students—rather like plant crops—are given pretests (the seedlings are weighed and measured) and then submitted to different experiences (treatment conditions). Subsequently, after a period of time, their attainment (growth or yield) is measured to indicate the relative deficiency of the methods (fertilizers) used."²⁵

At least two examples of the botanical model appeared in the literature during 1979. At DePauw University, Hardesty, Lovrich and Mannan carefully reported on the first-year evaluation of the school's Library Service Enhancement Program. Using control and experimental groups, together with pre- and post-testing, the investigators found that the students receiving instruction scored significantly higher, statistically significantly, than those in the control group on a 20-item test. Their mean average score rose from 12.2 correct items before instruction to 14.9 afterward. The test asked students to indicate which area of the library was "the most logical place to *start*" a search for information such as census data, a magazine article or a particular book.²⁶ Hardesty also measured the change in certain student attitudes before and after instruction, concluding that such change was much harder to effect than changes in library use skills. Finally, the article serves as the most explicit example to date of evaluation undertaken to justify bibliographic education to college and university administrators. While seeking

to explain clearly to other librarians the process of systematic evaluation and of creating a valid and reliable instrument, Hardesty noted that the evaluation efforts at DePauw "proved helpful in gaining administrative support" for the university's successful grant proposal for participation in the College Library Program sponsored by the National Endowment for the Humanities and the Council on Library Resources.²⁷

Other evaluations appeared in 1979 using tests which measure content. Phipps and Dickstein's description of their assessment of a library skills program at the University of Arizona is a useful complement to the DePauw study. Although not as helpful as Hardesty's description of certain evaluation procedures, Phipps and Dickstein went further in other respects, explaining how their evaluation was tied to explicit program objectives. They also showed scores for each question on their pretest and post-test, for both control and experimental groups, observing which library skills the program taught better than others, as well as which test questions proved more and less satisfactory.²⁸ Like the DePauw experimenters, Phipps and Dickstein concluded that bibliographic education significantly improved the library skills of the pupils studied, but they made no explicit claims of statistical significance or of impressing administrators.

From Penn State, Glogoff reported on the use of a homegrown test that was used, without modification, for a variety of disciplines in all classes receiving instruction. Since there was no attempt either to ascertain or demonstrate the value of instruction to students, no control group was established. The pretest at Penn State was used diagnostically to determine the level of library skills in a given class, and instruction for particular groups was modified accordingly.²⁹

The importance of using a control group when the purpose of evaluation is to assess the value of instruction versus no instruction was demonstrated in a study at Northeastern Oklahoma State University in 1977. Students enrolled in a library skills course did show improvement between pretest and post-test, but so did students not in the class. In fact, the scores of enrolled students showed no detectable difference from those of the other group. As a result, the librarians became aware that improvements in instructional methods were necessary.³⁰

A few years ago Wiggins reported testing at Brigham Young University. Programs designed to teach students how to use the card catalog and periodical indexes were developed with explicit objectives and were carefully evaluated to determine whether those objectives were attained. Test questions for both programs were piloted on students and librarians. Wiggins found that the scores of students receiving either pro-

Evaluating Bibliographic Education

grammed or nonprogrammed instruction showed statistically significant improvement on the post-test, while those of students in the control group did not.³¹

As the Wiggins studies illustrate, tests are employed not only as diagnostic tools or to measure the absorption of content, but also to compare methods of instruction—a time-honored subject of educational research.³² An activity gaining in popularity is the comparison of programmed and nonprogrammed instruction (the latter usually consists of one or more library lectures). As instruction librarians have become more aware of the great demands, actual or potential, on their time, they have often turned with hope to less personalized instructional methods.³³ Wiggins's studies, Surprenant's evaluation at Northland College of a program to teach use of the card catalog, and Phillips and Raup's treatment at Wooster Agricultural Technical Institute of their periodicals indexes program all used some version of a workbook and all found no loss of effect with programmed instruction. Wiggins even gave it a statistically significant edge over the library lecture.³⁴

Few studies, including a large number in the field of education, have documented the instructional superiority of one form of media over another.³⁵ A possible exception is Kuo's study, reported in 1973. A media librarian, Kuo reported on his comparison of six methods of instruction for the science library at Portland State College. He contrasted groups instructed through lecture, audio, slide-tape, notebooks with filmstrip, audiovisual (including a followup with a librarian), and a control group which received no instruction. An objective test of ninety items was given immediately following the various forms of instruction. Kuo concluded, not suprisingly, that the most effective format was the combination of audiovisual methods with a librarian-led session to answer questions and reinforce certain points.³⁶ Young in 1974 described Kuo's work as "the most elaborate experimental research on instructional strategies to date" in user education, a judgment that is still apt.³⁷ But, as Young observed then, the samples were small and some of the procedures involved in setting up the experiment were insufficiently explained.³⁸

Objective tests which attempt to determine the degree to which library skills are learned by students have certain advantages over other methods. They are relatively easy to administer and grade, and the results are readily quantifiable.³⁹ Tests share with other methods additional useful characteristics, such as pre-/post-administration and comparisons between or among groups. Yet, many thoughtful observers have raised serious questions about the widespread reliance upon tests

for evaluating user education. Kirk has even commented that "these tests...do not serve the profession very well."⁴⁰ Some of the criticism focuses on problems with methodology, such as lack of standardization or failure to establish validity and reliability of questions.⁴¹ "The greatest limitation, however," Young has noted, "to the prominent as well as to the numerous locally developed paper-and-pencil tests of library knowledge is their artificiality as devices for ascertaining a user's ability to negotiate the complex bibliographic structure of a library."⁴² Young voices a common complaint that, questions of methodology aside, tests can measure achievement of only the most fundamental user skills.⁴³ Two of the best tests, those at DePauw and Arizona, have focused on the most basic level, as their authors have readily acknowledged.⁴⁴

A central, and usually implicit, assumption of test makers for years has been that library usage would reflect whatever knowledge students could demonstrate on the objective tests, an assumption that is suspect.⁴⁵ Burton, while head of the instruction program at the University of Texas's Undergraduate Library, expressed it well: "True and false, multiple choice, and identification test items can measure whether students recall specific facts and principles about library materials and procedures; however, they cannot measure changes in behavior or actual success in finding material. Recall and behavior in a real library situation are not always analogous."⁴⁶ Others also stress that more effective library use is learned by actually using the library, and that testing for certain kinds of limited skills, often involving short-term recall, is not a satisfactory tool for evaluating bibliographic education. As Benson has noted: "User behavior must be our focus, not a prescribed set of skills....Do we care about differences in the ability to use libraries as a distinct issue? I am more concerned with the presumed goal underlying that of improved abilities: changes in the *actual* use of libraries."⁴⁷ Bloomfield has commented that, considering the narrow focus of library skills tests, "it appears that we librarians have shown a poor understanding of the value of the library to our students."⁴⁸ The artificiality imposed by the testing process itself has also not escaped comment, as critics have raised serious questions about the ability of testers to control the random and unpredictable variables that abound in the world of higher education.⁴⁹ Fjällbrant and Werking have even speculated that, combined with their narrowness, the attractiveness of objective tests has prompted instruction librarians to concentrate on basic, easily "measurable" instruction, and therefore has limited their efforts.⁵⁰

Evaluating Bibliographic Education

Another common complaint about objective tests is that students are often tested immediately after receiving instruction, and that the significance of such short-term gains is not likely to be great.⁵¹ The point is well taken but in all fairness, short-term evaluation is not a problem specific to objective tests; other assessment methods are also susceptible to its pitfalls. In an attempt to deal with this problem at Chalmers University in Sweden, Fjällbrant's evaluation procedures included an attempt to assess, through prestructured interviews, the long-term retention of library skills ten months after instruction.⁵²

More promising than objective tests are those measurement methods which go beyond the content of instruction to gauge student performance in the library. Kirk's well-known experiment in a large, introductory biology class at Earlham compared lecture-demonstration and guided exercise methods of instruction, finding no significant difference between the two groups. Instead of using only an objective test, Kirk had students write a short research paper—a product of their library use. Faculty evaluated the content, and Kirk assessed the bibliography on the basis of several criteria.⁵³ Similarly, as part of her Monteth College study, Knapp and her colleagues developed several performance tests that they concluded showed "considerable promise," but which needed more work.⁵⁴ Fjällbrant, too, in her multifaceted evaluation at Chalmers University, examined each student's list of references "in order to see whether the students were able to carry out a practical literature search."⁵⁵ Breivik's experiment at Brooklyn College compared library skills in two groups receiving instruction and a control group on the basis of grades received on research papers.⁵⁶

Yet the difficulty of constructing and administering performance evaluations is considerable, as indicated by their almost complete absence from the literature. Even Kirk and Knapp chose not to become involved with some features of objective testing which would have enhanced for others the value of their work, notably, the use of control groups and pretesting. In some cases, probably many, it seems to would-be evaluators "infeasible" and "too time-consuming" to administer performance tests to large numbers of students.⁵⁷ Hence, they rely on the easier objective tests, with all their drawbacks. Performance tests are probably much more common in separate library skills courses, when the entire course is under the control of librarians, than in course-related instruction.

The Survey

Although not treated widely in the literature, the survey of students

is the most common method of determining systematically the effects of bibliographic education.⁵⁸ Librarians use surveys to get feedback from students in order to improve instruction; to demonstrate to themselves, colleagues, faculty, and administrators the utility of their efforts; and to determine needs. By asking students whether instruction was worthwhile and, if so, how, they attempt to gauge its impact, as judged by the student, on both the process and the product of library use. Adams recently reported on the questionnaire used at SUNY-Oswego, which asked: (1) Did the course-related instruction help students complete their projects? (2) Was new information presented? (3) Did the students have problems completing the projects? (4) Were the presentation skills of librarians adequate? At Earlham, the survey evaluation of Kirk's experiment probably impressed faculty with the educational value of user education; three-fourths of the responding students in the introductory biology course thought they had learned as much or more about course content from the library-based paper assignments compared with the non-library-based exams and quizzes, and 95 percent said they had worked harder on the library-based papers.⁵⁹

Questionnaires usually ask students to compare instruction as presented with no instruction, rather than to compare two or three forms of bibliographic instruction. Response rates are usually quite good (with responses thus representative of the group), in course-related or separate-course instruction, assuming a small percentage of absences from class and assuming as well that the questionnaires are filled out and collected in class. Johnson, however, is leery of such terms as "useful" or "helpful," preferring more specific self-reporting on behavior in response to questions like, "How many times did you use *Biological Abstracts* during the last term?"⁶⁰ An issue in such cases, of course, is whether the student or the librarian is better able to determine the educational utility of the instruction. Johnson is assuming that the librarian must be, although even some librarians would disagree.

Another common use of questionnaires is to evaluate impersonal teaching mechanisms, usually "point-of-use" products such as computer-assisted instruction or audiovisual presentations. The programs often ask the user to fill out a nearby questionnaire and leave it at a designated spot.⁶¹ Not surprisingly, response rates to this appeal are usually quite low. At one institution, over a 14-month period, surveys of audiovisual programs received a response rate of between 4 and 6 percent. Undaunted, the evaluators went on to draw the conclusion that "while these are not sufficient response rates to be statistically significant, it can be said that response to the questions which attempted to

Evaluating Bibliographic Education

assess the value of this method of orientation indicated an almost total acceptance, in fact preference, for this method of instruction."⁶²

Finally, like objective tests, surveys are occasionally used to "pre-test," either diagnostically or as a way to determine the effect of instruction over a relatively long time. Frick at the University of Colorado-Colorado Springs and Werking at Lawrence University each conducted a survey to establish a baseline against which to measure progress when a similar survey was taken in the future.⁶³

Illuminative Evaluation

"Illuminative" evaluation has emerged among some instruction librarians in Europe, although there are as yet no reported cases of its use in the United States. This method deemphasizes the initial formulation of goals or objectives, and stresses instead participant observation and what Fjallbrant has termed "the expression of unexpected results.... Research is focused on what is actually happening in response to the innovation."⁶⁴ As a part of her evaluation at Chalmers University, students and faculty were interviewed about their participation in bibliographic education, but to what effect it is difficult to determine, given her sketchy description.⁶⁵ From Britain, Harris reported on the use of illuminative evaluation in an unusual research project, the Travelling Workshops Experiment of the Newcastle-Upon-Tyne Polytechnic Library. In that instance, evaluators wished to "produce insights rather than test hypotheses," and they relied heavily upon what Harris described as "subjective assessment," using chiefly observation and interviews.⁶⁶ It is too early to tell whether use of illuminative evaluation as such will grow to play a significant role in user education, but its disaffection for emphasizing quantitative methods seems to be part of an emerging trend among instruction librarians.⁶⁷

Statistical and Other Problems

It has long been fashionable to call upon librarians to learn about statistical methodology and research design, and evaluation of user education is one area that no doubt would profit from greater expertise and sensitivity. Some of the better evaluators have drawn with good effect upon the statistical knowledge of their colleagues in the local academic community.⁶⁸ Yet even some of these, as well as others, have been insufficiently judicious in their investigations or reporting. One recent evaluation sought to demonstrate that sixty-seven students who

on the pretest had responded to an attitudinal question at the lower (negative) two places of a five-point scale, showed "a strongly positive pattern of change" on the post-test. Actually, the majority of these registered either no change or a change in attitude toward the lowest point on the scale. No mention was made of the more than 100 respondents who had registered somewhere in the top three places on the scale; it would be interesting to learn whether as a group they had "improved" in attitude, stayed the same, or declined on the scale.⁶⁹ Also suspect are the returns from voluntary questionnaires; response rates are almost never given. One recent article mentions "a subjective questionnaire filled out voluntarily by students taking the workbook.... Questionnaire results were very positive."⁷⁰ The dangers of inference from such self-selected respondents are obvious. Another related problem is generalization on the basis of low response rates. One institution with a well-known user education program arrived at conclusions on the basis of responses to a questionnaire from 23 percent of 500 randomly selected seniors.⁷¹ To its credit, this library was apparently seeking to measure impact over a longer term than one semester or a few minutes.

Some of the problems relate to a fact now receiving greater attention: evaluation of user education is not only time-consuming, it is also a tough and tricky business.⁷² Psychologist Richard Johnson, speaking to a group of instruction librarians, made the point: "I can offer you no magic recipe to follow, no algorithm to learn, no ritual to perform which will insure that your instructional program will be automatically, adequately evaluated."⁷³ Indeed, psychologists and other social scientists are lacking such algorithms for the more general field of evaluation research. Cottrell has observed, "One gets the impression that what passes for evaluative research is indeed a mixed bag at best and chaos at worst," while Suchman agrees that the field "is notable for its lack of comparability and cumulateness of findings."⁷⁴ Knapp's report of the elaborate Monteith pilot project comments more than once on the difficulties of controlling variables such as different assignments, and students researching different topics and hence following different search strategies.⁷⁵ Moreover, as Suchman observes about evaluation generally, "the process...is highly complex and subjective."⁷⁶ At Texas, Burton received different advice about testing from six different departments on campus and concluded, "Research design is clearly as much an art as a science!"⁷⁷ Moreover, the teaching/learning process itself is very complex and subtle. Academics are hard-pressed to measure long-term gains in the mastery of more traditional subjects, such as philosophy and history. Perhaps some relief, substantive as well as

Evaluating Bibliographic Education

psychological, is to be derived from the recognition that because evaluation is so complex, no single tool or method can satisfactorily gauge a program's total effectiveness.⁷⁸

An even more fundamental problem for evaluators of user education is the lack of consensus about which library and bibliographic skills need to be transmitted to students, and the related disagreement on what to measure if success in the enterprise is to be determined. No single theme in the literature of evaluating bibliographic instruction is more pervasive than this complaint about the profession's lack of agreement on objectives. The absence of consensus is often significant not only within particular colleges and universities, but also within particular libraries, a point that has received insufficient attention.⁷⁹ Moreover, reports on evaluation from individual campuses often fail to state the program's objectives, and the reader must infer them. Yet the problem is much more acute at the national/professional level.⁸⁰ At a 1973 evaluation conference, Kirk declared: "The most important need in library instruction today is to have objectives. We simply do not have an adequate set of objectives."⁸¹ The absence of professional consensus on objectives means the absence of standardized tests or other forms of standardized evaluation.⁸² In his study of the library skills of future teachers, Perkins lamented that testing would be easier if librarians could agree upon "what knowledge is necessary to make full use of the resources contained in the library."⁸³ At the same time, he showed how remote such agreement is when he included among a "good" test's characteristics "that the individual taking the test finds it interesting and enjoyable so that he will cooperate," and when he asserted (correctly or not) that many librarians considered "an understanding of the Dewey Decimal Classification System...a necessity for intelligent library use."⁸⁴

Without standardized measuring tools and agreement on objectives, instruction librarians lack norms, whether for assessing a student's bibliography, answers on an objective test, or ratings tabulated from a questionnaire. At one institution the librarians concluded that because "the mean scores for Groups A and B indicated above 56 percent of the responses were correct, it is evident that both lectures and programmed methods have merit."⁸⁵ By the standards of many college professors, however, 60 percent is barely passing, and at another institution the students scored better than 60 percent on a pretest.⁸⁶ At Penn State the librarians chose 70 percent as the minimum score to indicate "satisfactory library skills," while at the University of Richmond 80 percent was passing.⁸⁷

On the other hand, the lack of norms at the national/professional level need not hinder local efforts. Professors in many disciplines have

for years been evaluating essay exams, research papers and bibliographies, master's theses, and other student work, all in the absence of national guidelines.

Questions of Proof

Several reasons for measuring the impact of bibliographic instruction involve justification of one kind or another, and there is considerable attention in the literature to "proving" the worth of instruction. Miller believes that "librarians are continually handicapped by the lack of substantive proof as to what library use instruction will really do for students," while Benson asserts more specifically that "proof must be in the form of aggregate statistical data, not individual anecdotes."⁸⁸ The recent appearance of several articles showing the use of objective tests to measure the impact of instruction is an attempt by librarians to provide more "objective" or "scientific" evidence, either of progress or of the relative value of different instructional strategies. The principal theme of Hardesty's article is captured in his quotation of a sentence from Suchman: "All social institutions or subsystems, whether medical, educational, religious, economic, or political, are required to provide 'proof' of their legitimacy and effectiveness in order to justify society's continued support."⁸⁹

Yet there is an important difference between demonstrating statistical significance and educational significance. One must ask whether the great concern with "proof," defined as statistical significance, is not much too narrow and perhaps counterproductive. It leads to measurement of the most basic levels of instruction, and may channel instruction itself in the same direction. While the administration at one college was sufficiently impressed by a program's gains (and certainly, it is far more important what the local community thinks than what professionals in other locales may prefer), an increase in the average number of correct test answers from 12.2 to 14.9 on a scale of 20 would not convince all administrations that bibliographic instruction was ipso facto worth the thousands of dollars it was consuming, whether the gain was statistically significant or not.⁹⁰ Brewer and Hills, among others, provide a healthy note of caution amid the calls for quantification by observing that the attempt to be *too* "scientific" has had drawbacks, most notably in the "universal adoption of evaluation strategies which are perhaps not those most fitted for the purpose."⁹¹ They refer rightly to "the complexity and subtlety of the teaching and learning process," and to the difficulty of measuring with overly narrow methodologies objec-

Evaluating Bibliographic Education

tives that are worth attaining.⁹² Parlett has criticized very well the “agricultural-botany paradigm”:

By imposing its own pattern and assumptions, the paradigm forces people to oversimplify, almost to the point of rendering the data meaningless. Conceived originally with massive samples and good controls in mind, it is nevertheless regarded as the model to be striven for, even if numbers are small; even if research situations are idiosyncratic in the extreme; even if random and uncontrollable factors intrude to a marked degree. Because it presents itself as objective, reliable, quantitative, and value-free, all—apparently—is forgiven.⁹³

Fortunately, some of the same individuals who call for “proof” use the term in a flexible manner. Miller believes that the well-regarded Earlham College instruction program “has proven that a course-related, sequential program of library-use instruction is feasible.”⁹⁴ Earlham has never undertaken to “prove” this quantitatively; it has satisfactorily demonstrated its worth, at home and abroad, and that is what Miller means. In a similar fashion, Miller counts the Eastern Michigan program a success, not on the basis of statistically significant evidence, but because the program has been incorporated into the university budget.⁹⁵ Even Suchman, upon whom Hardesty relies to emphasize the importance of “proof” of legitimate activities, is also flexible when it comes to what passes for proof. Continuing where Hardesty left off: “Both the demand for and the type of acceptable ‘proof’ will depend largely upon the nature of the relationship between the social institution and the public. In general, a balance will be struck between faith and fact.”⁹⁶

There are other dangers for instruction librarians in relying too much on statistical significance to prove the worth of a program, as demonstrated by the experience of sociologist George Conklin. Conklin helped develop an innovative introductory sociology course at a large private university in the Northeast. Although he was able to demonstrate statistically significant gains in learning by students, and although hitherto-declining enrollments grew from 155 to almost 300 within the year, the sociology department felt only lukewarm about the course, and significant opposition developed within the department to extending the innovations to another large course. Not long after, the department returned to the traditional method for handling large introductory classes, abandoning the innovations. “No one argued,” wrote Conklin, “that the teaching changes had not been effective, only that they were not needed for beginners anyway.” And he emphasized the importance of legitimizing innovation as a part of the academic prestige system.⁹⁷

Conclusion

The next few years should prove interesting for the evaluation of bibliographic education. For practitioners at the local level there is already much information available from which they can pick and choose to put together an instruction program, including evaluation procedures, designed to meet local needs. If they wish to gauge the impact of their program and perhaps also justify it to themselves and others, they will discover the variety of ways to go about it, that a spectrum of possibilities does exist.⁹⁸ No consensus will quickly emerge at the national/professional level about the goals of instruction or, therefore, about what or how to evaluate. Interest in objective tests for more basic instruction or orientation will probably continue. For both these reasons, the literature will continue to contain complaints about the lack of evaluation, although they should be fewer, and more of these will almost certainly specify the absence of evaluation for higher-level instruction.

There is some reason to hope that more studies will address that need. Because of the reports that have appeared in the 1970s, those librarians interested in using objective tests to evaluate a program or compare methods of instruction have much with which to work. It is reasonable to assume that some research may now more satisfactorily come to terms with library use and its relation to bibliographic education, turning from the agricultural-botany paradigm to what Parlett has called the "social anthropology paradigm."⁹⁹ Kirk and Knapp, in particular, have made good beginnings in evaluating performance, and instruction librarians should and probably will travel farther down that road. How much farther they will be able to go is an open question, considering the inherent difficulties noted here.

References

1. Brewer, James G., and Hills, P.J. "Evaluation of Reader Instruction," *Libri* 26: 56, 64, March 1976.
2. Lubans, John. Quoted in Fjällbrant, Nancy. "Evaluation in a User Education Programme," *Journal of Librarianship* 9:88, April 1977.
3. Benson, James. "Bibliographic Education: A Radical Assessment." In Cerise Oberman-Soroka, ed. *Proceedings from the Second Southeastern Conference on Approaches to Bibliographic Instruction, March 22-23, 1979*. Charleston, S.C., College of Charleston, 1980, pp. 53-68. I am indebted to Professor Benson for an early copy of his paper. See also Miller, Stuart W. "Library Use in Selected American Colleges." Master's thesis, University of Chicago, Graduate Library School, 1976, p. 16 (I am likewise

Evaluating Bibliographic Education

indebted to Mr. Miller for a copy of his thesis); Beeler, Richard J. "Introduction." In _____, ed. *Evaluating Library Use Instruction* (Library Orientation Series, no. 4). Ann Arbor, Mich., Pierian Press, 1975, p. vii; and Whitbeck, George W., and Hernon, Peter. "Bibliographic Instruction in Government Publications: Lecture Programs and Their Evaluation in American Academic Depository Libraries," *Government Publications Review* 4:1, 1977.

4. Hernon, Peter. "Library Lectures and Their Evaluation: A Survey," *Journal of Academic Librarianship* 1:16, July 1975. In a study devoted to documents instruction, Whitbeck and Hernon found that 80 percent of the programs were not collecting such data. See Whitbeck and Hernon, op. cit., p. 7.

5. Young, Arthur P., and Brennan, Exir B. "Bibliographic Instruction: A Review of Research and Applications." In John Lubans, Jr., ed. *Progress in Educating the Library User*. New York, Bowker, 1978, p. 19.

6. Kirk, Thomas G. "Bibliographic Instruction—A Review of Research." In Beeler, ed., op. cit., pp. 12-13.

7. Suchman, Edward A. *Evaluative Research: Principles and Practice in Public Service & Social Action Programs*. New York, Russell Sage Foundation, 1967, p. 11; and Lumsdaine, Arthur A. "Comments." In American Institutes for Research. *Evaluative Research: Strategies and Methods*. Pittsburgh, American Institutes for Research, 1970, p. 20.

8. Fjällbrant, op. cit., p. 84.

9. Lindgren, Jon. "Seeking a Useful Tradition for Library User Education in the College Library." In Lubans, ed., op. cit., pp. 82, 88.

10. Werking, Richard Hume. "The Place of Evaluation in Bibliographic Education." In Cerise Oberman-Soroka, ed. *Proceedings from the Southeastern Conference on Approaches to Bibliographic Instruction, March 16-17, 1978*. Charleston, S.C., College of Charleston, 1978, p. 100; Frick, Elizabeth. "Evaluating Student Knowledge of Facilities at the University of Colorado, Colorado Springs." In Carolyn A. Kirkendall, ed. *Improving Library Instruction: How to Teach and How to Evaluate* (Library Orientation Series, no. 9). Ann Arbor, Mich., Pierian Press, 1979, p. 101; Wiggins, Marvin E. "The Development of Library Use Instructional Programs," *College & Research Libraries* 33:475, Nov. 1972; and Phillips, Linda L., and Raup, E. Ann. "Comparing Methods for Teaching Use of Periodical Indexes," *Journal of Academic Librarianship* 4:423, Jan. 1979.

11. Kirk, op. cit., pp. 3-4.

12. _____. "Evaluation of Library Orientation and Instruction Programs: A Taxonomy." In Mary Bolner, ed. *Planning and Developing a Library Orientation Program* (Library Orientation Series, no. 3). Ann Arbor, Mich., Pierian Press, 1975, p. 47; Fjällbrant, op. cit., pp. 91-92; Vogel, J. Thomas. "A Critical Overview of the Evaluation of Library Instruction," *Drexel Library Quarterly* 8:322, July 1972; and Parlett, Malcolm. "Evaluation Innovations in Teaching." In H.J. Butcher and Ernest Rudd, eds. *Contemporary Problems in Higher Education: An Account of Research*. London, McGraw-Hill, 1972, p. 149.

13. For examples of such attitude surveys, see John Lubans, Jr. "Evaluating the Library-User Education Programs." In _____, ed. *Educating the Library User*. New York, Bowker, 1974, pp. 242-46; and Hardesty, Larry, et al. "Evaluating Library-Use Instruction," *College & Research Libraries* 40:309-17, July 1979.

14. Fjällbrant, op. cit., pp. 92-93; and Kirk, "Bibliographic Instruction," op. cit., pp. 3-4.

15. Adams, Mignon. "Effects of Evaluation on Teaching Methods." In Kirkendall, op. cit., p. 99.

16. Lindgren, op. cit., p. 88.

17. Werking, op. cit., p. 100; Kirk, "Bibliographic Instruction," op. cit., p. 12; and Farber, Evan I. "Library Instruction Throughout the Curriculum: Earlham College

- Program." In Lubans, *Educating*, op. cit., p. 159.
18. Johnson, Richard R. "Library Instruction: The Mythology of Evaluation." In Beeler, ed., op. cit., p. 36; and Lubans, "Evaluating," op. cit., p. 239.
19. Johnson, op. cit., p. 36.
20. Kirk, "Bibliographic Instruction," op. cit., p. 11.
21. Lindgren, op. cit., p. 88.
22. See Perkins, Ralph. "The Feagley Library Orientation Test for College Freshmen." In _____ . *The Prospective Teacher's Knowledge of Library Fundamentals*. New York, Scarecrow Press, 1965, pp. 153-93.
23. Bloomfield, Masse. "Testing for Library-Use Competence." In Lubans, *Educating*, op. cit., p. 221.
24. Barkeley, Margaret. "Arrows for Freshmen," *Library Journal* 64:402-04, May 15, 1939.
25. Parlett, M., and Hamilton, D. Quoted in Brewer and Hills, op. cit., p. 60. See also Parlett, op. cit., p. 146.
26. Hardesty, et al., op. cit., p. 317.
27. Ibid., p. 316.
28. Phipps, Shelley, and Dickstein, Ruth. "The Library Skills Program at the University of Arizona: Testing, Evaluation, & Critique," *Journal of Academic Librarianship* 5:204-14, Sept. 1979. In this respect, their article is a significant improvement over many which leave the reader guessing about how well students, before and after instruction, fared on particular questions.
29. Glogoff, Stuart. "Using Statistical Tests to Evaluate Library Instruction Sessions," *Journal of Academic Librarianship* 4:438-42, Jan. 1979.
30. Eymann, David H., and Nunley, Alven C., Jr. "The Effectiveness of Library Science 1011 in Teaching Bibliographic Skills." Tahlequah, Northeastern Oklahoma State University, 1977. (ED 150 962) The absence of a control group in the University of Richmond's evaluation weakens its case that freshmen improvement from a pretest to a post-test was a result of the Library Instruction Program. See DeLong, Edward J. "An Evaluative Report of the Richmond College Freshman Library Instruction Program." Richmond, Va., Richmond University, 1978, p. 24. (ED 157 547)
31. Wiggins, op. cit., pp. 473-79; and _____ . "Evaluation in the Instructional Psychology Model." In Beeler, ed., op. cit., p. 95.
32. Fjällbrant, op. cit., p. 88; Kirk, "Bibliographic Instruction," op. cit., p. 9; and Jamison, Dean, et al. "The Effectiveness of Alternative Instructional Media: A Survey," *Review of Educational Research* 44:1-67, Winter 1974.
33. Phillips and Raup, op. cit., pp. 420-21.
34. Wiggins, "Development," op. cit., p. 478; Surprenant, Thomas T. "A Comparison of Lecture and Programmed Instruction in the Teaching of Basic Catalog Card and Bibliographic Information—Results of a Pretest." In Oberman-Soroka, *Proceedings from the Southeastern Conference*, op. cit., p. 64; and Phillips and Raup, op. cit.
35. See Young and Brennan, op. cit., p. 23; and Jamison, et al., op. cit., p. 55.
36. Kuo, Frank F. "A Comparison of Six Versions of Science Library Instruction," *College & Research Libraries* 34:287-90, July 1973.
37. Young, Arthur P. "Research on Library-User Education: A Review Essay." In Lubans, *Educating*, op. cit., p. 9.
38. For additional criticism, see Brewer and Hills, op. cit., p. 57.
39. Burton, Susan. "Objective Tests as an Evaluation Tool: Problems in Construction and Use." In Hannelore B. Rader, ed. *Library Instruction in the Seventies: State of the Art* (Library Orientation Series, no. 7). Ann Arbor, Mich., Pierian Press, 1977, p. 100.
40. Kirk, "Evaluation," op. cit., p. 44.
41. Burton, op. cit., p. 100; and Young, op. cit., p. 7. Not all tests are equally vulnerable to this criticism. More than other writers, Hardesty took considerable care to explain how he and his colleagues established reliability for their test at DePauw. See Hardesty, et al., op. cit., pp. 311-12.

Evaluating Bibliographic Education

42. Young, op. cit., p. 7.
43. See Knapp, Patricia B. *The Monteith College Library Experiment*. New York, Scarecrow Press, 1966, p. 71; and Fjällbrant, op. cit., p. 89.
44. Phipps and Dickstein, op. cit., p. 206; and Hardesty, et al., op. cit., p. 310.
45. Kirk, "Evaluation," op. cit., p. 44; and Gebhard, Patricia. "How to Evaluate Library Instructional Programs," *California Librarian* 37:38, April 1976. Surprenant is one of the few experimenters who has made this assumption explicit. See Surprenant, op. cit., p. 58.
46. Burton, op. cit., p. 100.
47. Benson, op. cit., pp. 15, 9. See also Knapp, Patricia B. "The Methodology and Results of the Monteith Pilot Project," *Library Trends* 13:85, July 1964; and Bogel, op. cit., p. 316.
48. Bloomfield, op. cit., p. 229.
49. Brewer and Hills, op. cit., p. 57; and Fjällbrant, op. cit., p. 86.
50. Fjällbrant, op. cit., pp. 89-90; and Werking, op. cit., p. 109.
51. Kirk, "Bibliographic Instruction," op. cit., p. 10; Burton, op. cit., p. 101; and Benson, op. cit., p. 9.
52. Fjällbrant, op. cit., pp. 91-92.
53. Kirk, Thomas G. "A Comparison of Two Methods of Library Instruction for Students in Introductory Biology," *College & Research Libraries* 32:465-74, Nov. 1971; and Wiggins, "Development," op. cit.
54. Knapp, *Monteith*, op. cit., p. 72.
55. Fjällbrant, op. cit., p. 91. See also Dummer, E. Heyse. "The Library Method Course in the Postwar Era," *AAUP Bulletin* 32:329, Summer 1946.
56. Breivik, Patricia Senn. *Open Admissions and the Academic Library*. Chicago, ALA, 1977; and _____ . "Effects of Library-Based Instruction on the Academic Success of Disadvantaged Undergraduates." In Hannelore B. Rader, ed. *Academic Library Instruction; Objectives, Programs, and Faculty Involvement* (Library Orientation Series, no. 5). Ann Arbor, Mich., Pierian Press, 1975, pp. 45-55.
57. Keever, Ellen H., and Raymond, James C. "Integrated Library Instruction on the University Campus: Experiment at the University of Alabama," *Journal of Academic Librarianship* 2:187, Sept. 1976; and Eyman and Nunley, op. cit., p. 10.
58. Lindgren, op. cit., p. 88.
59. Adams, op. cit., pp. 99-100; and Kirk, "Comparison," op. cit., pp. 471-72.
60. Johnson, op. cit., p. 35.
61. Stevens, Charles H., and Gardner, Jeffrey J. "Point-of-Use Library Instruction." In Lubans, *Educating*, op. cit., pp. 273-74; Lubans, "Evaluating," op. cit., pp. 247-48, 252-53; Hacker, Betty, and Stevens, Richard. "Evaluating the AIMLO Project." In Beeler, ed., op. cit., p. 53; and Davis, Elisabeth B., et al. "A Two-Phased Model for Library Instruction," *Bulletin of the Medical Library Association* 65:44, Jan. 1977.
62. Hacker and Stevens, op. cit., p. 53.
63. Frick, op. cit., p. 101; and Werking, op. cit., pp. 107, 111-17. See also Paterson, Ellen R. "An Assessment of College Student Library Skills," *RQ* 17:226-29, Spring 1978, for a useful variation. Paterson used as a diagnostic pretest an open-ended question, asking students what library sources they would use to compile a bibliography on certain subjects. Such a method would probably be worthwhile in a pretest/post-test arrangement seeking to judge the impact of instruction.
64. Fjällbrant, op. cit., p. 86.
65. *Ibid.*, p. 92.
66. Harris, Colin. "Illuminative Evaluation of User Education Programmes," *Aslib Proceedings* 29:354, Oct. 1977.
67. *Ibid.*; and Fjällbrant, op. cit., p. 86.
68. Hardesty, et al., op. cit., p. 311; Wiggins, "Development," op. cit., p. 473; and Phipps and Dickstein, op. cit., p. 213.
69. Hardesty, et al., op. cit., pp. 314, 315.

RICHARD HUME WERKING

70. Phipps and Dickstein, op. cit., pp. 206, 214.
71. Miller, op. cit., p. 60.
72. Brewer and Hills, op. cit., pp. 63-64.
73. Johnson, op. cit., p. 32.
74. Cottrell, Leonard S., Jr. "Foreword." *In* Suchman, op. cit., p. viii; and Suchman, op. cit., p. 27.
75. Knapp, *Monteith*, op. cit., pp. 147, 152.
76. Suchman, op. cit., p. 11.
77. Burton, op. cit., p. 101.
78. Glogoff, op. cit., p. 430; and Fjällbrant, op. cit., p. 87.
79. Werking, op. cit., pp. 101, 109. The problem is by no means unique to librarians. According to Parlett, "it is rare indeed for teachers even to approach agreement on objectives" (Parlett, op. cit., p. 147).
80. *See, for example*, Bloomfield, op. cit., pp. 226, 230; Lubans, "Evaluating," op. cit., p. 232; Farber, op. cit., pp. 159-60; Miller, op. cit., p. 83; Perkins, op. cit., p. 50; Hernon, op. cit., pp. 17-18; Oakley, Adeline D. Content Analysis of Student Responses in Topic-Centered Library Orientation. Ann Arbor, Mich., University Microfilms, 1979, p. 83; and Whilden, Sara L. "Planning a Library Instruction Program Based on Measurable Objectives." *In* Rader, *Academic Library Instruction*, op. cit., pp. 13-21.
81. Kirk, "Bibliographic Instruction," op. cit., pp. 10-11. The "guidelines" formulated in 1975* by the ACRL Bibliographic Instruction Task Force are an attempt in this direction, but they have yet to serve directly as the basis of reported evaluation.
82. Burton, op. cit., pp. 101-02.
83. Perkins, op. cit., p. 50.
84. *Ibid.*, pp. 48, 52.
85. Phillips and Raup, op. cit., p. 423. *See also* Whilden, op. cit., p. 17.
86. Hardesty, et al., op. cit., p. 313.
87. Glogoff, op. cit., p. 441; and DeLong, op. cit., p. 13.
88. Miller, op. cit., pp. 78-79; and Benson, op. cit., pp. 10-11.
89. Hardesty, et al., op. cit., p. 309; and Suchman, op. cit., p. 2.
90. Hardesty, et al., op. cit.
91. Brewer and Hills, op. cit., p. 57.
92. *Ibid.*, pp. 60-61.
93. Parlett, op. cit., pp. 148-49. *See also* Fjällbrant, op. cit., p. 87; Schumbacher, Mark. "Letters to the Editor," *Journal of Academic Librarianship* 5:92-94, May 1979; and Adams, Mignon. "Letters to the Editor," *Journal of Academic Librarianship* 5:94, May 1979.
94. Miller, op. cit., p. 58.
95. *Ibid.*, p. 61.
96. Suchman, op. cit., p. 2.
97. Conklin, George H. "The Frustration of Academic Innovation: Can New Teaching Techniques Survive in the Academic Marketplace?" *Teaching Sociology* 5:125-40, Jan. 1978.
98. Brewer and Hills, op. cit., p. 60.
99. Parlett, op. cit., pp. 150-53.

This Page Intentionally Left Blank