Problem Solving Through Experimental Research: The Need for Better Controls

SHARON L. BAKER

ABSTRACT

GOOD LIBRARIANS ARE ALWAYS "experimenting"—that is, trying something new (generally to solve some library problem) and watching to see if it works. Unfortunately, since they generally fail to establish appropriate experimental controls, librarians are often left with rather subjective impressions of whether or not their new ideas have worked.

This article discusses the types of controls which need to be established in experimental studies to ensure that the conclusions reached are valid, using actual library experiments to illustrate the points made. Questions that librarians need to ask when they are trying to determine if it is worthwhile investing time and money in experimental research are also suggested.

INTRODUCTION

An experiment is designed to test a hypothesis—i.e., a tentative generalization concerning the relationship between two or more variables in some situation (Mouly, 1978). Examples of simple hypotheses are: variable *X* causes variable *Y* to occur and "installing a series of signs within a community college library will decrease the number of directional reference questions asked by patrons." The most popular form of an experiment is the standard pretest-posttest experimental design. For example, in a college library, Joe, who is head of reference services, might count, during a three-month pretest period, the number of directional reference questions asked by patrons. Joe would then install direction signs and count the number of directional reference questions asked during a second three-month period. Finally, he would compare

Sharon L. Baker, School of Library and Information Science, 3072 Library, University of Iowa, Iowa City, IA 52242
LIBRARY TRENDS, Vol. 38, No. 2, Fall 1989, pp. 204-14

the number of questions asked before and after the installation of signs to see whether a significant change occurred. If a change had occurred, Joe might be able to conclude that installing the signs influenced the number of questions asked if he has taken precautions to see that the experiment has internal validity.

Internal Validity and Experimental Studies

An internally valid study is one in which the researcher has tried to ensure that it is variable *X*—rather than some other variable—that caused *Y* to occur. The researcher does this by establishing a series of experimental controls to eliminate any compounding variable or form of bias that could influence study results.

The writer's first experiment, conducted with two lab partners for a college science class, was not internally valid. Dale grasped a different object in each hand. Arms extended at a uniform height in front of him, he tried to release the objects at exactly the same instant. My job was to stoop near the floor and record which object, in each of ten pairs of objects, hit first. Cheryl videotaped the process. We spent twenty minutes dropping and recording, dropping and recording, carefully noting minute differences in the "hit" rate, certain that we were disproving the hypothesis that Isaac Newton developed so long ago: the force of gravity causes all objects to accelerate toward earth at the same rate (32 feet per second squared).

We were, of course, wrong. When the instructor slowed down the videotape, advancing it frame by frame, we could see that the small differences in the time it took for each object to hit the floor were caused by a series of human errors. It was impossible for Dale to hold two objects at exactly the same distance from the floor and drop them at precisely the same instant; it was impossible to gauge exactly when each one hit the floor. Because these compounding factors were not controlled for, the conclusions were suspect.

All too often, librarians working in the field test their new ideas without establishing appropriate controls. As a result, their conclusions may be invalid. Last year a medium-sized public library in the United States (name withheld upon request) discovered that it answered 50 percent of its reference questions accurately—i.e., completely and correctly. To remedy this problem, the director organized a series of workshops for the staff on conducting better reference interviews and on developing more systematic search strategies. When the accuracy rate increased to 70 percent, the director concluded that the workshops had caused the improvement.

This experiment has a major problem that is not readily apparent—i.e., the director failed to realize that some factor other than the experimental treatment applied (that is, the workshop sessions) might have caused accuracy to increase. A ten-minute conversation with this library administrator identified four other possible causes of the

improvement: between the pretest and the posttest measurements of accuracy, the worst reference librarian on the staff was fired, the library strengthened its reference collection substantially, and the number of staff working on the reference desk at any one time was significantly increased thus allowing individual librarians to spend a longer time answering individual patron questions. Also, the fourth possible cause for improvement was that questions asked during the two measurement periods were not scrutinized to make certain that they were of equivalent difficulty. It is, therefore, impossible to say conclusively that the reference workshops led to improvement in accuracy.

Even trained investigators find it difficult to conduct, in complex social science settings, experiments that are internally valid because they lack the controlled facilities to study the effects of a change in isolation. That is, they are not working in environments in which rival explanations are easily ruled out. It is not as difficult to conduct internally valid studies in the physical sciences. If the research department of a major fertilizer company wants to determine whether a newly developed fertilizer strain works better than its five major competitors, the scientists involved can begin with six similar sets of corn seed. The scientists can hold all other factors constant in the company's lab—e.g., giving each set of seeds identical amounts of water and light, placing them in soil with identical composition, keeping the growing temperature equivalent, etc. In a controlled experiment of this type, if the seeds treated with the new fertilizer grow an average of 20 percent more than each of the other sets of seeds, there is a reason to believe that the fertilizer caused the greater growth.

Now consider a "simple" library experiment. A school library notices that its 16mm films receive little use. The librarian decides to promote them, going around to individual teachers and hawking the films that relate to each person's teaching area. It is fairly easy for her to measure use before and after the promotion. She may be able to show that use increased and that the increase came after the promotion was made, two necessary conditions for proving causality. But it will be impossible for her to hold constant all other factors that might affect use, especially those that affect individual teachers. For example, a teacher who feels ill may show a film rather than lecture. Use of a particular title might increase because of a renewed interest in that subject area—witness last year's resurgence of interest in John F. Kennedy on the 25th anniversary of his assassination. Teachers could be made aware of films through some means other than the librarian's promotion (e.g., recommendation by a friend). Or interest in films could be renewed due to budget cuts that curtail videotape rental. Determining whether the librarian's promotion has caused the circulation increase is difficult in this complex environment.

Still, trained researchers can identify most, if not all, of the compounding factors that can affect experimental results. They control for

as many factors as they can, then closely monitor the environment to determine if still other factors could be influencing study results.

Examples of Internally Valid Studies

Described later are two studies, conducted in public libraries, that established fairly rigorous experimental controls. In the first, Harris and Michell (1986) tried to determine whether library patrons who observed a competent reference interview would be significantly influenced by the "social context" of the interaction. The researchers explored six factors that would possibly affect patron ratings of the librarian's behavior:

1. gender of the librarian;
2. gender of the patron asking the reference question;
3. gender of the person observing the transaction;
4. sex-role identity of the person observing the transaction (the degree to which the observer thinks of himself/herself as primarily masculine, primarily feminine, or as a person who exhibits both masculine and feminine characteristics);
5. warmth of the librarian (the degree to which the librarian smiles, maintains eye contact with the patron, has a friendly tone of voice, and exhibits open body posture); and
6. the level of inclusion exhibited by the librarian (the degree to which the librarian instructs patrons in the use of reference tools).

Rather than having observers watch any reference interview, the researchers developed a series of sixteen videotapes in which four of the experimental factors were purposely and systematically varied—the gender of the librarian, the gender of the patron, the level of warmth displayed by the librarian, and the level of inclusion displayed by the librarian. Other controls used in the study were as follows:

1. To avoid bias due to the question itself, all patrons were asked the same question on a noncontroversial topic (ridding houseplants of insects). In each of the sixteen videotapes, the librarian gave the patron a complete and correct answer to the question.
2. Professional actors, all of whom received instruction from the same professional drama coach, played the parts of the patron and the librarian so that problems with acting ability would not affect final results.
3. To avoid bias from having different words used in each reference interview, only two scripts were used. One illustrated high inclusion on the part of the librarian and one low inclusion. Both scripts involved a multistep reference process.
4. Students enrolled in a graduate library science program prescreened the tapes. Without knowing the purpose of the study, they verified that each tape exhibited the behavior it was meant to illustrate—that is, high or low inclusion, high or low warmth. The students also

verified that the behavior of the male and female librarians on the tapes was similar.
5. A research assistant approached equal numbers of males and females in the lobby of a medium-sized public library and asked them if they would be willing to participate. To avoid bias, observers were randomly assigned to watch one of the sixteen videotapes.
6. Observers were not told the true purpose of the study since that too could have affected study results. Instead, they were asked if they would be willing to participate in a study about "reference work in libraries."
7. The test used to measure the observer's sex-role identity was one developed by a psychologist; it had previously been found to be a valid and reliable way of measuring this quality.

Harris and Michell (1986) found that observers rated "warm" librarians significantly better than "nonwarm" ones. Female observers felt that librarians displaying low inclusion were warmer and more professional while male observers interpreted librarians displaying high inclusion as warm.

The second study utilizing a series of experimental controls was one this writer conducted after reviewing several studies showing that book displays increase the use of the titles they contain (Baker, 1986). The experiment was designed to determine why displays increase use. Two hypotheses were tested. The first was that fiction titles that are displayed in prime locations (where they are highly visible and accessible to browsers) will circulate significantly more than their counterparts on the regular shelves or in a nonprime display location. The second hypothesis sought to test whether displays worked because they narrowed readers' choices by guiding them to a small collection of titles, thus overcoming the effects of information overload. (Information overload is the confusion and indecision which patrons can feel when they are confronted with too many choices from which to make their selections. The potential for overload appears to be very great in all except small libraries.) The study attempted to see whether one narrowing strategy, that of recommendation, would cause fiction titles marked "recommended" to circulate more than their counterparts no matter where they were located—i.e., in a prime display area, a nonprime display area, or on the regular shelves.

The following experimental controls were used:

1. To make sure that factors relating to one library were not influencing any circulation changes, identical studies were conducted in two unrelated libraries located thirty miles apart. The libraries differed in their collection size and also in the service philosophy of staff.
2. To avoid bias in the selection process, books were randomly chosen from the fiction collections in each library and were randomly assigned to six different treatment groups: (a) prime display, recom-

mendation; (b) prime display, no recommendation; (c) nonprime display, recommendation; (d) nonprime display, no recommendation; (e) regular shelves, recommendation; and (f) regular shelves, no recommendation. Group *e* titles, which were not displayed, served as a control for the display half of the experiment. Group *f* titles, which remained on the regular shelves and received no recommendation, acted as a control for the recommendation half of the experiment.

3. Books in each treatment group were compared and found to be similar to each other and representative of the general fiction collection in terms of other variables thought to affect use—i.e., the book's age, its length, its past circulation history, its physical condition, its format (paperback or hardback), and its cover (with or without jacket).
4. All books in each treatment group were left in their regular locations for a three-month pretest period, and circulation was measured. The experimental treatment was then applied and circulation was remeasured for a three-month posttest period.
5. The signs used to promote the prime display books and the carts used to house them were identical to those used for books on the nonprime display. Staff did not promote any study titles during the course of the experiment, and a weekly shelving check ensured that study titles were reshelved in the appropriate location.
6. To provide further control, the variables of location and recommendation were reversed during a third, three-month posttest period, and circulation was remeasured. That is, the books that had been displayed prominently during the initial posttest period were moved to the nonprime display and vice versa; books with no recommendation during the initial period were marked "recommended" and vice versa. The reversal of experimental treatments on the same books was done to ensure that any circulation increases were due to the experimental treatments rather than caused by any unique qualities of the books themselves.
7. Circulation was graphed on a week-by-week basis for the entire nine months to verify that circulation changes corresponded with application of the experimental treatments.
8. A statistical test controlled for the normal, seasonal variations in circulation that occurred in both libraries. Circulation of a random group of fiction titles during the entire nine months of the study was also observed to verify that no other factor was affecting overall use of the fiction collection.
9. Patrons who checked out books in the six treatment groups were interviewed to determine why they had selected those particular titles.
10. Finally, the type of experimental design chosen for the study, the standard pretest-posttest model, controlled for the effects of eight technical factors that could have affected the experiment's internal

validity: history, maturation, testing, instrumentation, statistical regression, experimental mortality, selection bias, and selection-maturation interaction. (A further explanation of these threats to internal validity and controls in experimental design can be found in Campbell and Stanley [1963].)

Study results showed that prime display locations significantly increased the use of the titles involved, but nonprime locations did not. Recommended books were used significantly more than nonrecommended books in the larger of the two test libraries where readers experienced more information overload.

Even though neither of these experiments was conducted in a fixed laboratory environment, the researchers identified and controlled for a number of variables that might have influenced the results. As such, both experiments had internal validity—they were as unbiased as possible, ruling out rival explanations for their findings by controlling the environments in which the studies were conducted as much as possible. In each case, it was reasonable to conclude, in the libraries in which the studies were conducted, that the experimental treatments influenced the behavior of the observers or patrons.

External Validity and Experimental Studies

In laboratory settings, external validity is a secondary goal. That is, scientists attempt to establish controls rigorous enough to enable them to generalize the findings beyond one particular setting. Then they can show that variable *X* will always cause variable *Y* to occur and can state scientific laws, such as the law of gravity, which are always true. Social science researchers, on the other hand, are dealing with a more complex environment—i.e., the field of human behavior. Humans are such complicated beings that literally dozens of variables may influence us to behave in certain ways. As a result, it is difficult, if not impossible, for social science researchers to attain external validity (Guba & Lincoln, 1981; Krathwohl, 1985), or to come up with fixed laws stating principles that should be followed in every situation. Rather, social science researchers end up developing more complicated theories—theories which state that variable *Y* will change (or occur) under certain conditions but not under others. While the absence of fixed laws governing human behavior makes the job of the library researcher more difficult, it also makes it more fascinating.

Like reference librarians, experimental researchers must be puzzle-solvers. They must be able to determine why one treatment caused reaction *A* in one library and reaction *B* in another. Consider the following example. Over the past decade, seven studies examined the use of booklists in libraries. In each case the researcher measured initial circulation of a set of titles then remeasured circulation after promoting titles through a booklist. Circulation of booklist titles remained sub-

stantially the same in three of the studies. Circulation increased significantly in the other four.

Each study was internally valid. That is, within the test library environment the researcher established controls for various factors that might have affected use of the booklist titles in that particular library environment including characteristics relating to the books chosen, the subject of the list, the presence or absence of annotations for each title, and seasonal variations in circulation. But it was not possible for the researchers to control for all variables influencing use.

What factor caused circulation to increase in some cases but not in others? This puzzle can be solved by examining facets of each study to see if one explanation will account for the differences in results. In this case, the factor that appears to have caused booklist titles to circulate in some libraries but not in others is the method of the list's promotion.

Elsewhere, this writer argues that, in order for a particular title to be used, large numbers of patrons have to become aware of that title and have to feel that it will meet some personal need for recreation or enlightenment (Baker, 1986a). Many more people will be exposed to a work than will ever actually want to use it, in the same way that many more shoppers will notice oatmeal on the grocery shelves than will ever actually buy it. And patrons prefer to use works that they find convenient to obtain. If libraries want use of certain titles to increase, librarians should design promotional methods that meet two criteria—that they are easy to use, and that they expose large numbers of patrons to specific titles.

In three of the seven studies, the promotion method did not significantly increase use of booklist titles because it violated these two principles. Taylor (1982), following the test library's usual practice, left booklists out for voluntary patron pickup in a number of unobtrusive spots in the library. As a result, few patrons noticed the lists, picked them up, or used them in their selection. In the other two studies, booklists were not distributed within the library where they would be easy for patrons to use. Rather, they were, in one case, stuffed in student mailboxes at a university (Powell, 1972) and, in the other case, shown on cable television commercials (Auld, 1978). The potential for reaching a fairly large number of patrons with specific reading needs was there, but the convenience factor was missing; persons who wanted the titles were not at the library when they recognized the need. They had to save the lists and then take them to the library at some later date.

Those libraries that increased use of booklist titles did not violate these promotion principles. Lists were promoted within the library in such a way that many patrons saw them. Goldhor (1981) and Golden (1983) gave one to each adult patron, Parrish (1986) displayed the lists prominently at the entrance of the library, and Wood (1985) gave them to patrons who were having difficulty finding fiction titles classified within the Library of Congress scheme.

This example illustrates that researchers have to work harder to come up with theories that will apply to many different situations that exist in complex environments like libraries. Experimental research can, if properly designed by competent researchers, help build these types of theories.

When Should Experimental Research be Conducted?

A related issue is how often practicing librarians should test the workability of their new ideas through experimental research (or for that matter, through survey, historical, or other types of research). The answer to this question will differ from library to library and from situation to situation. Generally, a librarian considering the possibility of researching some issue should ask the following questions:

1. Does the change have the potential to greatly improve service or to save a large sum of money for the library over time?
2. Are staff members available and willing to conduct such research? Or, alternately, can the library afford to hire a consultant to design and implement a study?
3. Does the person assigned to conduct the research have the type of specialized training that is needed to design, conduct, and analyze research studies? (This training includes, at a minimum, an in-depth knowledge of the principles of hypothesis testing, causality, study design, and statistics.)
4. Is the library willing to endure the inconvenience of some of the special controls that will need to be established? For example, staff members may be required to keep special statistics, to change their behavior for a short period of time, or to ask patrons to fill out questionnaires to provide supportive evidence for the hypothesis being tested.
5. Is the library willing to bear the expense of the experimental study? While some studies are relatively cheap (costing only a few thousand dollars to design and implement), others may be quite expensive.
6. Is the library willing to use results of the experimental study to make changes? Management must be willing to use experimental results to make the changes indicated. Staff must be willing to drop ideas that do not work, as well as to adopt new strategies shown to be effective.

Only if the answer to each of these questions is a resounding "yes," should a library consider conducting the type of rigorously controlled studies discussed here. Particular attention must be paid to ensuring that appropriate personnel are assigned to or hired for the research project—personnel who are competently trained in the complex and complicated business of research design, implementation, and analysis. Managers are failing their responsibilities if the problems identified are inadequately investigated or if solutions based on inadequate investigations are allowed to be made into practice (Allen, 1986).

The project described later, for example, might benefit from this type of experimental research. Suppose a large public library has established that its community has literacy needs that are not being met by another organization. The library is considering devoting $100,000 for each of the next ten years to a literacy program that will train tutors to work with adult illiterates and will provide both the materials and the space needed for the tutoring. This extensive monetary investment makes it worthwhile for the library to hire a researcher, preferably before the program is instituted, who can establish, from the beginning, controls that will enable the library to see whether its program will be successful. Over the course of the first year, the researcher might try to determine:

1. which promotional method(s) attracts the most potential tutors;
2. which promotional method(s) successfully attracts illiterate adults to the library's program;
3. which tutoring method (e.g., the Laubach method) has the best results in teaching people how to read; and
4. what practices the library can follow to decrease the dropout rate among both the tutors and illiterate adults.

Over the long run, determining these points should save the library money even if the initial outlay for the researcher's efforts costs $10,000 or more.

A few libraries, such as the Fairfax County (Virginia) Public Library, have established offices of research, statistics, or evaluation to help them solve important problems of this nature, while others have established "visiting researcher" positions for this purpose. Still other libraries have banded together to solve common problems inexpensively. The Library Research Center at the University of Illinois persuaded libraries to donate $1,000 each to explore a common problem—finding a quick and accurate way to measure in-house use (Rubin, 1986). Three public libraries in Virginia designed and carried out a joint project to determine: (1) how accurately they answered reference questions, and (2) how they could increase the accuracy rate (Rodger & Goodwin, 1984).

Summary

Using the guidelines given earlier, an individual library can determine whether it is willing to invest the time and money in an experimental study designed to solve a specific library problem. The key to successful experimentation is a highly trained, competent researcher who establishes rigorous controls to attain internal validity. That is, the researcher shows that in this one library, under these specified conditions, a change in variable Y resulted from a change in variable X, rather than from a change in variable Q or A. The staff at this library benefits directly from the experiment using study findings to solve the initial problem.

While results of an experiment conducted in a single library cannot be generalized to other libraries, they will, if published for professional scrutiny, benefit other librarians indirectly. When scholars in any discipline accumulate a large body of research on a single topic, they begin to notice consistent patterns. That is, they begin to see that variable *X* will change variable *Y* in certain situations but not in others. This intense scrutiny eventually leads researchers to develop theories, like the one previously described on information overload, that can predict human behavior to some extent. Such theories can guide daily operations in libraries of different sizes and types, benefiting the profession as a whole.

References

Allen, G. G. (1986). Management and the conduct of in-house library research. *Library and Information Science Research, 8*(April/June), 155-162.

Auld, L. W. S. (1978). *The effect on public library circulation of advertising via cable television.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Baker, S. L. (1986a). The display phenomenon: An exploration into factors causing the increased circulation of displayed books. *Library Quarterly, 56*(July), 237-257.

Baker, S. L. (1986b). Overload, browsers and selections. *Library and Information Science Research, 8*(October/December), 315-329.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago, IL: Rand McNally and Company.

Golden, G. A. (1983). *Motivation to select books: A study of annotated and unannotated booklists in public libraries.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Goldhor, H. (1981) Experimental effects on the choice of books borrowed by public library adult patrons. *Library Quarterly, 51*(July), 253-268.

Guba, E. G., & Lincoln, Y. S.(1981). *Effective evaluation.* San Francisco, CA: Jossey-Bass.

Harris, R. M., & Michell, B. G. (1986). The social context of reference work: Assessing the effects of gender and communication skills on observers' judgements of competence. *Library and Information Science Research, 8*(January/March), 85-101.

Krathwohl, D. R. (1985). *Social and behavioral science research.* San Francisco, CA: Jossey-Bass.

Mouly, G. J. (1978). *Educational research: The art and science of investigation.* Boston, MA: Allyn and Bacon.

Parrish, N. B. (1986). *The effect of a booklist on the circulation of fiction books which have not been borrowed from a public library in four years or longer.* Unpublished master's thesis, University of North Carolina at Greensboro.

Powell, R. (1972). *The effect of a booklist on library circulation.* Unpublished manuscript, University of Illinois at Urbana-Champaign.

Rodger, E. J., & Goodwin, J. (1984). *Reference accuracy at the Fairfax County public library,* Washington, D.C.: Metropolitan Washington Library Council.

Rubin, R. (1986). *In-house use of materials in public libraries.* Urbana-Champaign, IL: University of Illinois, Graduate School of Library and Information Science.

Taylor, M. T. (1982). *The effect of bibliographic accessibility upon physical accessibility of materials in a public library setting.* Unpublished doctoral dissertation, University of Michigan.

Wood, R. J. (1985). The experimental effects of fiction booklists on circulation in an academic library. *RQ, 24*(Summer), 427-432.