Evaluating Your Own Program

Stephen F. Hamilton

If you can't afford an outside evaluator, you can do your own evaluation. Be concerned both with "Does it work?" and "How does it work?"

The demand for accountability often takes the form of requirements that new programs be evaluated. Teachers and others who have ideas and wish to try out new approaches in education, therefore, must be prepared to respond to sometimes hostile questions about a special program both at the planning stage and after an initial trial period. Those questions are usually a variation on, "Does it work?"

The innovator could, of course, respond, "Does standing up in front of a classroom work?" Although it is true that educational research has not answered some basic questions about conventional practices and it is likely that many of those practices would not withstand careful evaluation, only the new approach faces demands for evaluation. Pointing this out seldom reduces the demands.

The irony continues. While professors of education and social science with their degrees and six-figure grants struggle to demonstrate that one educational practice has different effects from another, the innovative practitioner is asked to prove the value of his or her program while directing its implementation. Even if funds are allocated especially for evaluation, which rarely happens, they are never sufficient to interest the Rand Corporation in bidding for the job.

Most of the growing literature on evaluation is written for full-time researchers with extensive training. The educational innovator usually has to design an evaluation without much professional advice, conduct it with limited resources, and defend it without the advantage of comparisons to conventional approaches that have been similarly evaluated. The following suggestions will not eliminate these problems, but they should help educators conduct evaluations of programs they direct.

From Informal to Formal Evaluation

One of the first hurdles to jump is the belief that evaluation is a magic art that produces hidden truths if only you know the right formula. A good evaluator can, indeed, uncover unsuspected information and provide keen insights by using sophisticated data collection and analysis. But the program director evaluating his or her own program must use techniques that are already available or easily accessible. These simpler methods can also be revealing, but they must be used systematically and thoughtfully.

The place to begin is with the informal means you already use to evaluate a program. All educators make judgments about the value of programs they read about, observe, or participate in. These are based on evaluation procedures that may be good or bad, explicit or unexamined. They must be brought to the surface and examined; then the best ones can be made more formal and augmented so they can serve to inform others instead of remaining private. That, after all, is the chief difference between a formal evaluation and the kind that we all use informally. A formal evaluation is public and must communicate effectively to others.

The need to communicate relates to the question of audience (for whom is an evaluation being performed?) but that question is logically preceded by the question of what an evaluation is focused on and what purposes it serves.

The question, "Does it work?" is the most obvious focus for an evaluation. A second hurdle for evaluators is the assumption that this question is the only legitimate one to try to answer. "How does it work?" is also an important question. The two, outcome and process, should both be addressed.

Dealing with the outcome question, "Does it work?" requires statement of some objectives. The rewards of planning a program using measurable objectives are great at this point.[1] If such objectives have not been prepared during the program planning process, they must be before the evaluation of outcomes

[1] Robert F. Mager, *Preparing Instructional Objectives* (Palo Alto, Calif.: Fearon Publishers, 1962) is a standard source for this purpose. *Planning By Objectives*, a manual published by ACTION's National Student Volunteer Program, 806 Connecticut Avenue, N.W., Washington, D.C. 20525, is very useful and is distributed free.

can proceed. However, the use of objectives should not blind the evaluator to unanticipated outcomes. Provisions should be made for detecting and gathering evidence about both desired and undesirable outcomes that were not expected.

The process question, "How does it work?" has to do with the vast area between objectives and outcomes. What actually occurred in the program? Who was involved? What did they do? In what ways did the program depart from wishes and plans? Ideally, sufficient information about process will be reported to allow replication or adaptation of the program by others.

Purposes and Audiences

The dual focus on both outcome and process allows an evaluation to serve multiple purposes. It may help the evaluator to think of all of those purposes as being related to decision making about the program. The information that is collected and reported should be useful to people who are making decisions.

The most common purpose for evaluation is accountability. The evaluation provides a justification for the expenditure of resources on a program. Obviously, "Does it work?" is a key question for this purpose, but "How does it work?" is equally important because it includes the cost factor in the cost/benefit calculation. A program that succeeds in raising the reading levels of five normal children by one grade level over a year's time at a cost of $100,000 has worked but is not justifiable. And monetary costs are not the only critical concern. A program that offends parents or burns out teachers may not be desirable even if it is effective.

The first audience for accountability evaluation should be the program staff. They, after all, are the ones who should be most concerned about the expenditure of their time and talents. Often, however, program staff are willing to act on the basis of their informal, private evaluations and see no need for formal evaluation. In fact, they are likely to quarrel with any evaluation that conflicts with their common sense. One advantage to evaluating your own program is that it obviates conflict between an outside evaluator and program staff.

But program sponsors are the most frequent consumers of evaluation for accountability. Principals, superintendents, boards of education, and "the public" want to know whether their energy, their reputations, and their tax dollars should continue to be devoted to a new educational program. The challenge to program staff members is to collect and present the kind of evidence that will convince these outsiders that their program is worthwhile. This does not place the staff members in the position of unbiased searchers after truth, but if they are responsible and honest in their efforts, their work will be legitimate and useful.

A second purpose for evaluation is improving the program. Outcome and process evaluation, both during and at the conclusion of a program, can provide information upon which decisions can be based to continue or alter certain practices. The chief audience for evaluation with this purpose is program staff and, sometimes, participants. Program sponsors may be another. They are sometimes more favorably disposed toward continuing a program that has identified both its strengths and its weaknesses. One means of improving a program using evaluation information is by revising objectives. Some objectives may prove to be unrealistic and others too easily obtainable. Other, previously unstated, objectives may come to the fore as the program progresses. Restating objectives can improve efficiency, effectiveness, and morale, and it can result in a more favorable showing in terms of accountability.

A third purpose for evaluation is dissemination. Other people may share the objectives and values underlying your project. If they learn of positive outcomes they will be interested in the process by which those outcomes were obtained. A detailed description and thoughtful analysis of the way the program took shape and operated is most valuable for other educators wishing to adopt or adapt the program. For this audience, the frank inclusion of dead ends, wrong turns, fortunate coincidences, and other unflattering information is invaluable. You need not spare the self-congratulation, but they will admire you more if your report helps them replicate your program than if you make yourself sound superhuman.

Timing

Concern for evaluation should begin during the planning stages of a program. One of the reasons it is often difficult to prove that a program is worthwhile is that evaluation is left until the program has ended. By that time it may be too late to gather needed evidence. Furthermore, evaluation after a program has ended has no value in improving the program while it is in operation. During the planning stages, attention should be given to what decisions must be made at what time, and what audiences will be interested in evaluation reports. The evaluation schedule and specific procedures should be planned to provide relevant information to each audience in time to make necessary decisions.

During the implementation and maintenance stages, the major focus of evaluation should be on process, but in order to assess process, some indicators

of outcomes are needed. Since the primary audience for evaluation at this stage is the program supervisor and participants, measures of outcome need not be formal but they should be thoughtful. Participants' expressed feelings that things are not going well enough may be sufficient evidence that changes are needed. However, it is best to rely on more than one indicator. Often it is helpful to decide on some indicators or standards in advance. For example, having a certain number of active participants at a certain stage of the program may be set as an indicator that the program is proceeding satisfactorily. This procedure clearly identifies needed changes when standards are not met, and when they are, it can provide welcome reassurance to program leaders.

The program's original objectives should be compared to accomplishments. If program participants are different from the stated target group, it may be necessary to change recruiting procedures. It may also be necessary to change objectives. If one of the original objectives of the program was to teach students about the structure and functions of both local and state government, but the participants have become fascinated by local government and wish to explore that level more fully, then explicit changes in objectives will make the program more comprehensible to outsiders and may point the way to some significant changes in procedure. For example, you may want to name the local family court judge to the advisory committee instead of the state senator. Significant changes in objectives might require renegotiation with a sponsor too. If the League of Women Voters is helping with the program, they will need to be informed about such a change in objectives or they might consider the program a failure for never getting to the state level.

During the concluding stage of a program, whether it is the end of the program or just the end of a year, outcome evaluation for outside audiences becomes more important. This is the time when evidence is needed to inform outsiders' judgments about the program's value. It is also the time to address other people who might want to try a similar program. Include in the final report or annual report enough information about process so that other people can avoid your mistakes and capitalize on your successes.

Documentation and Evaluation

It is useful to distinguish between documentation and evaluation. Documentation is the collection of available information and its recording in usable form. Evaluation involves making judgments, often on the basis of documentary evidence. It is an interpretive process. Many times requests for evaluation are in fact only requests for documentation. For example, a

basic fact about any educational program is the number of participants. Recording and reporting that number is documentation. Assessing the match between the actual number of participants and the desired number is evaluation.

The first step in an evaluation is careful documentation so that information will be accessible. Documentation requires recording and collecting. Keep records of what you do and copies of relevant materials. Letters, news releases, fliers, materials used in the program, and materials prepared by participants should all be collected for future reference. Your regular record of activities or a special journal devoted to the program will also be valuable records. You should plan to record anything pertinent that cannot be easily obtained if you need it later.

You should also keep records on participants, ideally including some information about individual characteristics such as age, racial or ethnic group, social class, and pertinent background information.

Visual documentation is important. Photographs can enliven presentations, publications, and exhibits. Take pictures during all stages of the program. Take them yourself, have participants take them, or bring in a photographer. Often participants or parents will be glad to take pictures. Videotape and eight millimeter film are other visual media that are relatively simple to use and inexpensive. Before and during the program, think about how you can show people who weren't involved what the program was about. Make sure you have raw materials when it comes time to make your report.

Outcome Evaluation: Design and Methods

The place to begin evaluating program outcomes is with the outcome objectives from your proposal or plan. If your objectives are stated in behavioral terms, criteria for judging whether the outcomes are satisfactory are included, and you have only to demonstrate that those criteria have been met to establish that your program was effective. Evidence of participants learning from a program is difficult to establish directly. Often, use of a special instrument is necessary. The three most commonly used types of instruments are tests, questionnaires, and interviews.

Teachers are very familiar with one kind of evaluation, the evaluation of individual performance–grading. You cannot convince anyone that your program was a good one by reporting that every student earned an A, but you can use some of the same data used in assessing students' individual achievements to demonstrate the value of the program as a whole. The difference is that the "unit of analysis," the entity to be evaluated, is not the student but the group of program participants.

Teacher-made tests based on the content of lessons are the most familiar source of data for individual evaluation. Reporting scores or the mean score on such tests is no more useful in program evaluation than reporting final grades, but if the content of those tests and the performance standards are reported, the skeptic might be convinced that students had indeed learned some important things from the program. Such a presentation is more effective if pretest scores are available on the same or an equivalent test given at the beginning of the program. Otherwise, the skeptic can argue that the students knew all of that before they entered the program and could have done as well without participating.

Administering the same test to a control group of students who are not in the program can deal with that objection too, but creating a control group in the real world is extremely difficult. You must admit students to your program on a random basis and convince those who were not selected to take the test. A less rigorous nonrandom comparison group can help, but even that is difficult to obtain. Will another teacher administer your test to his or her class in order to help you show your approach is more effective? How can you demonstrate that the program participants and comparison group are the same except for the program? Matching is the usual method, and it is not terribly good. What characteristics should be matched? Sex, year in school, and IQ are probably important. So are grade point averages, race, and socioeconomic status, parents' education and occupation. The list goes on but it is already too long to be practical without a huge sample and a computer. Control and comparison groups are potentially of tremendous value in educational evaluation and should be used whenever possible. They are unlikely to be used in evaluations conducted by practitioners.

Standardized tests can be seen as providing a built-in comparison group—all of those students whose scores determined the test norms. If program participants gain two grade levels in a year's time, there is a basis for claiming that the program was effective. When the same standardized tests are administered to the entire school, the claim can be strengthened if other students in the same school did not do so well. A within-school comparison is necessary if program participants represent a group that usually performs better than the norms: white, middle class, suburban students, for example.

The trade-off in using standardized tests is that they may not tap the special learning taking place in an innovative program. Their use is most safely limited to demonstrating that those involved in the program did not fall behind the conventional classes in their performance on standard measures of academic achievement. For programs that involve radical de-

partures from conventional practices, such as extensive time out of the classroom, this may be very important.

Criterion-referenced tests are particularly appropriate for use in program evaluation. The best known standardized tests are norm-referenced. Scores are established in comparison with norms so that the test-takers are always compared with each other. Criterion-referenced test scores are stated in terms of the proportion of correct responses made. If 80 percent of program participants can answer 80 percent or more questions on a tough test, that is powerful evidence that the program worked.

Although paper-and-pencil tests of cognitive knowledge and skills are the most common types of tests, there are others that may be even more useful. There are attitude measures that may capture important program effects of a different order. It is also possible to establish performance tests as demonstrations of learning; music teachers do this all the time. If students are expected to learn about city planning, they may be given a planning task to do and be evaluated on their performance at the task. The use of expert judges is often helpful in making performance tests work. If two or three professional city planners critique the students' work, the tests will be more challenging and interesting to the students and more convincing to outsiders as a demonstration of program effectiveness.

Questionnaires are generally used to assess attitudes rather than knowledge. "Closed-ended" questionnaires are easiest to process. In these all likely responses are listed and respondents simply check the one that best represents their opinion. As anyone who has filled out such a questionnaire knows, this can be frustrating because it is never possible to think of all reasonable responses in advance. "Open-ended" questionnaires avoid this problem by allowing respondents to write in their own words. The trade-off is that such responses are much more difficult to code and analyze and that some respondents may have difficulty expressing themselves in writing. A good way to develop a questionnaire is to ask participants to write or dictate statements about the program. These statements can then be edited and placed in a questionnaire format so that respondents can check whether they agree strongly, agree, have no opinion, disagree, or disagree strongly. This procedure is best done quickly so that results can be reported to the participants and they can talk as a group about what the results mean. This discussion then becomes another source of data for use in the evaluation.[2]

Interviews may also be either closed-ended or open-ended. The advantage of an interview over a questionnaire is that the interviewer can probe for further comments, clarification, and explanation of statements. In addition, respondents usually speak more easily than they write. The trade-off is that interviews, particularly open-ended interviews, take tremendous amounts of time to transcribe, code, and analyze. It is better, when needed resources are not available, to make a point of talking informally with participants and keeping careful notes on their responses to questions asked in those informal situations. Exact quotes are good to have whenever possible because they can add interest and concreteness to evaluation reports.

Group interviews are an alternative to interviews of individuals. This procedure involves recording the statements of participants in a group discussion. As with individual interviews, tape recording is ideal, but careful notes will do. Group interviews have the advantage of revealing participants' group process skills, which may be an important objective. They are, however, more difficult to conduct and to analyze than individual interviews.

Tests, questionnaires, and interviews can be administered to all participants or to just a sample. If a sample is used, be sure to report how it was selected. Any of these three types of instruments may be administered before, during, and after participation in a program to indicate changes in participants. When selecting an instrument it is very important to be clear whether you are interested in participants' knowledge, their skills, or their attitudes. The type of instrument used and its content will vary considerably according to what it is intended to measure.

Some additional sources of evidence on program outcomes are: "unobtrusive measures," journals or written reports by participants, testimonials, and observations.

"Unobtrusive measures"[3] or indicators can be very convincing because they are not influenced by your biases. Two good examples were given by the supervisor of a program intended to introduce youth to local government. He noted that after young people began attending school board meetings, cocoa was served at the meetings along with the usual tea or coffee. He also reported that a committee of the county legislature changed a meeting time for the convenience of a student who was observing the meetings. Both of these changes in procedures to accommodate youth were indicators that the young people were being taken seriously by adult members of governing bodies.

Journals or written reports by participants can

[2] This procedure is described along with other issues related to participant uses of evaluation in Robert M. Rippey, ed., *Studies in Transactional Evaluation* (Berkeley, Calif.: McCutchan, 1973).

[3] This term comes from a book by E. J. Webb, D. T. Campbell, R. D. Schwartz, and L. Sechrest, *Unobtrusive Measures: Nonreactive Research in the Social Sciences* (Chicago: Rand McNally, 1966).

contain valuable information about participants' experiences in a program and can reflect some of their learning. If participants are required to do this kind of writing, they must know in advance what is expected and understand its purpose. Ideally participants would be involved in deciding on such requirements.

Testimonials can come from leading citizens, other teachers, parents, or anyone else who has a basis for assessing a program's outcomes and who has credibility with the evaluation's audience. Testimonials might be solicited through letters, conversations, questionnaires, or interviews, or they might be unsolicited. Exact quotes are useful, but be sure to get permission to attribute a statement to a specific individual.

Observations may be so obvious as to be ignored. What will make your observations as a program leader credible to an outside audience is specification of behavior. If you say, "Keith really gained a lot from this," your observation will have little impact. But your observation can be very good evidence if you are prepared to say, "When Keith came to his first meeting, he hardly said two words. After the group decided to do an ecological study of the river, he not only made appointments and talked to several experts, but he also arranged with a science teacher to instruct the group in water testing procedures. Now the other students recognize Keith as a leader and respect him for his knowledge and ideas."

Process Evaluation

The examples of evaluation methods described so far all relate to outcomes and are particularly appropriate for the purpose of justifying a program at the end of the year. It is just as important to evaluate process and to keep track of objectives while the program is in progress. Evaluation of process depends heavily on documentation. The judgments enter when you try to decide what inputs were essential, what steps were crucial, and what factors really made a difference. You might describe a very logical set of ten steps that you followed in organizing a program, but you might have reason to believe that what really made the program go was the enthusiasm and leadership given by three participants. In this case it would be more useful for your own purposes and certainly for others interested in trying your approach to pay attention to how these leaders were found and what motivated them.

Process evaluation can make use of most of the same methods used in outcome evaluation—especially interviews, questionnaires, unobtrusive measures, writings by participants, and observations—but the audience and timing may be different. The questionnaire constructed from participants' statements is particularly appropriate. An important use for process evaluation is in making decisions about a program while it is in operation. For this purpose the audience includes program leaders and participants, and results must be known very quickly. Process evaluation is also useful in making decisions about continuing or modifying a program. It is extremely useful to other people who would like to try a similar program.

Stephen F. Hamilton is Assistant Professor, Department of Human Development and Family Studies, Cornell University, Ithaca, New York.

How to Begin

A careful evaluation requires considerable time, energy, and thought. Obviously the resources devoted to evaluation must be proportional to those available to the program as a whole. A good evaluation is well worth the effort because it strengthens the program.

In planning an evaluation for a particular program, you should begin by deciding how the program's specific objectives can be measured. (General objectives may not be measurable.) Next you should decide what purposes the evaluation should serve and what audiences it will have. This involves identifying decisions that should be based on the evaluation and noting when those decisions must be made. Then appropriate methods can be selected. Finally, a schedule should be made and specific responsibilities assigned so that it is clear what evaluation activities are going to occur at what times and who will be responsible for them.

The two most important questions to keep in mind when planning an evaluation are: "What do I want to learn?" and "What will I do with what I learn?" If you can answer these two questions, you have a good start. EL

Recommended Reading

A useful reference aimed at practitioners rather than researchers is *Evaluating Service-Learning Programs: A Guide for Program Coordinators*, ACTION pamphlet No. 4300.7 (4/78), which is a companion volume to *Planning By Objectives* and available free from the same source. The following are more technical but may provide helpful background:

Caro, Francis G., ed. *Readings in Evaluation Research.* New York: Russell Sage Foundation, 1971; Weiss, Carol H. *Evaluation Research: Methods of Assessing Program Effectiveness.* Englewood Cliffs, N.J.: Prentice-Hall, 1972; Patton, Michael Q. *Utilization-Focused Evaluation.* Beverly Hills, Calif.: Sage Publications, 1978.