Designing Program Evaluations: A Circular Model

Robert J. Menges

Northwestern University

Marilla Svinicki

University of Texas, Austin

Sooner or later, every faculty and instructional development program is faced with the need for evaluation, be it evaluation of a single workshop or of the program as a whole. Most developers have little formal training in program evaluation and approach this task with some trepidation and uncertainty. If their exposure to program evaluation is confined to the single point of view most characteristic of research in their own discipline, they are understandably limited in the evaluation approaches they consider applicable.

This article outlines an approach to program evaluation that can allow even novice evaluators to produce creative and thorough designs. This approach combines the techniques of creative thinking with research methodology from a variety of specializations. We believe it to be logical and commonsensical, a realistic approach to designing program evaluations.

Traditional Approaches

In early writings about evaluation and even in some contemporary texts, the recommended approach is that of the rationalist. According to Tyler's classic statement, evaluation is "the process of determining to what extent the educational objectives are actually being realized" (1950, P. 69).

The three major activities included in the rationalist approach are based on the scientific method. They are: (a) The evaluator frames questions which the evaluation is to answer; (b) Data appropriate to the questions are identified and collected; and (c) Answers to each question are formulated as the data permit. This highly linear approach to evaluation might be depicted as follows:

Generate questions	Identify and collect data	Answer questions
-----------------------	------------------------------	---------------------

More recently, an alternative to the rationalist approach has become popular. Usually termed "naturalistic," this model assumes that evaluation questions emerge from the data. Instead of beginning with questions to be answered or objectives to be verified, this approach begins with events to be documented. These events are examined in the context where they naturally occur.

As a chief advocate of the naturalistic approach, Guba emphasizes that it is "a paradigm for inquiry, not a method" (1981, p. 76). It is characterized by certain key assumptions, each of which is quite different from the assumptions made by the rationalist. For example, the naturalistic paradigm assumes that there are multiple realities (rather than a single reality to be discovered), and that findings are more likely to diverge than to converge as an inquiry proceeds. With regard to research design, naturalists assume that the design is emergent, rather than preordinate; that is, they assume that the design will change over time as it incorporates the experience of both investigators and respondents.

A linear depiction of the naturalistic approach looks like this:

Identify and collect data	Generate questions	Answer questions
------------------------------	-----------------------	---------------------

We believe that, in reality, most program evaluations reflect both approaches and that they are considerably more complex than either approach in its stark linear form. The usefulness of an evaluation is likely to be enhanced if we recognize that the essential character of evaluation is circular rather than linear. Generating questions neither necessarily precedes nor follows data identification; these activities occur in simultaneous interaction. Choice of method for data collection and of time for data collection are also interdependent. Data analysis and interpretation are not sequential; they are parallel, and they inform one another. Figure 1 portrays these relationships in what we call the circular model of evaluation.

The Circular Model of Evaluation

The circular model, as shown in Figure 1, has three parts: one circle depicts the design phase, a second circle depicts the execution phase, and a third circle links the first two.

As we can see by looking at the top circle, we first discuss what questions the evaluation will ask and what sources of information contribute to identification of those questions (the *what* issue). Next, we deal with the sources of data, that is, *whom* to ask, and then with occasions for data collection, that is, *when* to ask. Finally, we discuss the various forms information may take, that is, the issue of *how* data are collected.

This paper deals only with the design phase, but the model would not be complete without two more circles. The lower circle, the execution phase, includes actual collection, analysis, and interpretation of data. The upper and lower circles are linked by a third circle, illustrating the interdependence of planning and implementation. Not only does planning influence data collection, analysis, and interpretation, but the reverse is also true. Preliminary data analyses may lead to revised questions or they may suggest new forms of data to be contributed by other sources and gathered on different occasions.

FIGURE 1. The Circular Model

Generating Questions: Deciding What to Ask

Just as a circle has no beginning or end, this model has no beginning or end. We must start somewhere, however, so we begin with potential sources of questions.

Identifying question sources.

The left column of Figure 2 lists potential sources of questions for the evaluation. Most of these sources are also potential audiences for answers to the questions. For example, those who design the program can contribute specific questions and will also be very interested in the answers. Some program goals are probably articulated in documents that initiated the program, but others may be unstated and best articulated by those who designed the program.

Other sources noted in Figure 2 include program participants, who may be students and instructors in innovative classes or clients who use a new service. Non-participant observers are outsiders hired specifically to

Source	Question
Program Designers	Were the goals met? Was the program efficient?
Program Participants	Is this transferable to my situation? What did I learn? Did I like it? Is it for me?
Non-participant Observers	How reliable are the other data sources? Were there non-content issues which should be considered?
Leaders/Facilitators	How effective was I? Which activities were most effective? How much did the participants learn?
Consumers	How effective was the program at producing changes? How useful are the changes to me?
Documents	What are the stated goals? How do they differ from the achieved goals? Who are the consumers? What is the target audience? How efficient was the program?
Other	E.g. audiotapes - How effective were any presentations? Who said what?

FIGURE 2. Identifying Questions and Question Sources: Deciding What to Ask

provide “objective” information about the program. They could be interviewers of participants, classroom observers, external panelists who judge program outcomes, and so on. Leaders/facilitators are those actually running the program. They might have designed the program, although functioning in a leader role gives them new concerns and a different perspective at the time data are collected; or they might be workshop leaders other than the program designers. Consumers of the program include teachers and students, administrators, parents, library users, teaching assistant supervisors, business people who hire program graduates, and so on. What these consumers have in common is that they are affected by changes the program induces in participants and by decisions made as a result of the program.

“Documents” refers to items that might have helped to initiate the project or to materials that were generated during program planning. Proposals and reports that led to establishing the program might contain explicit or implied statements of purpose, which could become questions for the evaluation to answer. Existing examples of student work from old programs might provide clues to questions that should be asked of new programs. Logs of daily activities kept during program planning might reveal concerns that never made it into the final proposal, but that might be important for evaluators to consider.

Identifying questions.

Not all of these sources are appropriate for all programs; nor do they exhaust potential question sources. They are listed to stimulate evaluators’ thoughts about who might have vested interests in the program and about where there might be pertinent documentation. Once a list of potential question sources is laid out, the evaluator frames questions of interest, resulting in questions like those in the right column of Figure 2. These sample questions are generic and might apply to almost any program. In the example below, we suggest specific questions that might be generated during this step.

The next step is to select the actual questions to address. The potential questions just mentioned may suggest patterns to the evaluator and thus lead to specific questions for the evaluation. Other questions may arise from only one source; whether those questions are included depends on the importance of that source to the project as a whole. If the question is about cost, for example, and if it is raised by those who will eventually have to fund the program, the topic deserves serious consideration. Answers

may determine whether the program will continue beyond its experimental stage.

Identifying Data Sources: Deciding Whom to Ask

Figure 3 suggests which source can contribute data appropriate for each question. For each cell of this matrix, the evaluator tries to generate two or three possibilities, regardless of feasibility. The initial purpose is to identify possibilities that might otherwise be overlooked. Some cells will make no sense or will be useless and can be skipped. But before a cell is written off, its possibilities should be explored. What may at first appear to be impossible or useless can sometimes yield very interesting insights.

Once multiple alternatives are generated, the evaluator can go back and review cells to determine which data would be most informative as

Data Source

Question		Designers	Participants	Observers	Leaders	Consumers	Documents	Other
A	Were the program goals met?	x	x	x	x	x		
B	How efficient was the program?	x	x	?	x		x	
C	Can the skills be transferred to another setting?		x	x		x	x	

etc.

Alternative ways of generating data for Question B

Source

Designers	quality of product; training needed for leaders
Participants	time logs of work required; scaled question about time required
Observers	timed observations; number/type of questions asked
Leaders	time/effort for preparation; number of questions/mistakes
Consumers	quality of product; "sense" of job being rushed
Documents	cost sheets; time sheets; work orders for support
Other	

X's represent reasonable alternatives.

FIGURE 3. Identifying Data Sources: Deciding Whom to Ask

cross-checks, which would be most economical and feasible, and which would be most reliable and valid. Some combinations will make more sense or seem more desirable, such as cross-checks between teachers and students, supervisors and employees, designers and archival documents. Other considerations include time and effort constraints for any given data source. For example, one does not want the evaluation to overwhelm participants to the detriment of their involvement in the program. Non-participant observers might require too much training. Cost is always a factor in deciding how much use can be made of non-participant observers; these costs must be minimized.

Occasions for Data Collection: Deciding When to Ask

Data from a particular source may change across time as perspectives change with experience. New questions might occur as the evaluator contemplates what is revealed by each data source at different points in a project. Initially, one might consider asking only about the extent to which

Sample Question: What goals are being accomplished?

Who	When			
	Before	During	Immediately after	After a delay
Designer	List of goals		Narrative evaluation of outcomes	
Participants	Free response List of goals as see them	Self-report form of progress	Questionnaire; performance measure	Interview or questionnaire
Observer		Narrative Checklist Slice of life	Summary evaluation	
Leader		Log of progress Checklist	Summary evaluation	
Consumer				Questionnaire on effectiveness of participants
Documents	List of goals	Products related to goals	Action plan analysis	Follow-up on action plan

FIGURE 4. Occasions for Data Collection: Deciding When to Ask

a training process is useful in producing a skill. By evaluating across time, one can also determine whether skill acquisition is a continuous process or a discrete, all-or-nothing occurrence.

To guide decisions about when to collect data, a matrix is produced for each question being asked (Figure 4). If the question under consideration is "What goals are being accomplished by this program?" participants might be asked before the program about their expectations and goals, perhaps as part of the registration form. During the program they might be asked to identify which goals seem to fit with each activity, using a scaled self-report form. Immediately after the program, they might be given a questionnaire and asked to assess their level of confidence that the program goals have been met. After a delay, they might be interviewed by phone or surveyed by mail to assess the appropriateness of the goals for their everyday job experience.

Methods for Collecting Data: Deciding How to Ask

The process of choosing data collection techniques can stimulate creative planning and can offer possibilities that might not otherwise be considered, provided the evaluator is willing to mix evaluation approaches. The variety of methods can be arrayed along a continuum. At one end are methods favored by naturalistic researchers. These methods are "emic," that is, data are expressed in the categories and meanings of the respondent. With these measures, the subject freely responds as he or she sees fit. Emic methods include narratives, logs, journals, open interviews, slice-of-life tapes, and self-generated cases. Evaluators subsequently extract from these data the ideas and incidents that have some bearing on questions under study. Analysis may also yield additional questions for subsequent evaluation.

At the other end of the continuum are structured methods of data collection familiar to most researchers. These "etic" methods use the researcher's meanings and categories and impose them on data. Yielding readily quantifiable data, etic methods include pre-structured questionnaires, checklists, rating scales, behavior coding for observational data, and so on.

Between these extremes are methods that allow some flexibility by accommodating the richness of respondent-generated data while also facilitating evaluator-oriented analysis. These methods include structured interviews, guided free-response logs where participants respond in their

own words to pre-set questions, and structured case studies where fundamental questions are established by the evaluator.

Tapping a variety of data collection methods makes the evaluation more interesting and perhaps easier. Since several sources provide data for a question, the actual measurement process can be less rigorous than if there were only one source. The variety of methods can stimulate evaluators to gather data on questions and issues that a less open system might preclude.

Putting it All Together

Figure 5 serves as a convenient summary. Each question to be answered is matched with its data source according to the time when data are collected. The chosen data collection method is then entered in that cell of the figure. Armed with this action plan, program evaluators may begin to implement the evaluation.

		Occasion						
		After a delay						
		Immediately after						
		During						
		Before						
Source	Designer							
	Participants							
	Observers							
	Leaders							
	Consumers							
	Documents							
	Others							
		A	B	C	D	E	F	G
		Question						

FIGURE 5. Choosing Data Collection Method According to Question, Source and Occasion

But the design process does not stop here. Even as the program is in progress, and the evaluation is underway, evaluators must be alert to new questions and data sources that present themselves during execution of the plan. For example, participants may spontaneously engage in program development by requesting new activities or supplemental materials not envisioned in the original design. Or observers may become privy to a "behind-the-scenes" information exchange which would be of value in assessing the program, but could not have been anticipated. Evaluators should be open to this more "generative" aspect of design. Analysis and interpretation, the reflection phase of evaluation that is usually left until the conclusion of the program, should be an ongoing activity. New categories of data sources and questions should be allowed to emerge during the data collection so that design and execution occur simultaneously once the program begins. In Figure 1, the design phase and the execution phase are linked, illustrating these interactions.

A Faculty Development Example

Our application of the circular model is a composite drawn from evaluations we have conducted. The activity to be evaluated is a summer institute for university faculty from several colleges in a state. The program was initiated at the behest of the state coordinating board and aims to provide renewal opportunities for faculty from smaller public institutions that do not have resources for extensive faculty development programs of their own. The board hopes that the opportunity to spend two weeks at the state's flagship institution taking intensive work from prominent faculty in residence there will upgrade instructors' content knowledge and their teaching repertoires, in turn producing renewed commitment and improved performance on their home campuses. The board also hopes that the collegial atmosphere of the institute will stimulate cooperative programs and resource-sharing between institutions.

The board requires both host and invited institutions to contribute toward costs of the institute. Participants' home institutions provide a stipend and travel costs. The host institution provides staff and faculty time as well as access to campus libraries and recreational facilities. The state board provides housing expenses and materials as well as administrative costs. Faculty from participating institutions apply to the program and submit a work plan, discussing how their participation would benefit the home campus and their own professional growth.

To see how this program might be evaluated, we proceed through the circular model.

Identifying Questions and Question Sources

Potential sources of questions and of data include the state board, the host institution, its participating faculty and staff, the invited institutions and their participating faculty, the peers and students of participants, non-participant observers, and relevant documents. From this list we can brainstorm questions that might be of interest, as we did in Figure 2. For example, students at invited institutions would be able to answer questions about the classroom teaching skills of participating faculty. Colleagues of participating faculty would be interested in how participation affects teaching and how it influences content expertise. The host institution and the state board would be interested in how difficult it was to organize and implement the program, who participated, what the institution got out of hosting, how much the program cost, and so on.

Various documents might serve as data sources: for example, the evaluator could read about the original program goals in the project proposals, could learn about faculty goals and expectations by examining application forms, and could obtain information about the results of participation from action plans prepared by participants at the end of the institute. In addition to these questions, evaluators would want to address the traditional concerns of participants, such as the match between program goals and individual goals, how much each learned that was useful, and whether the program was enjoyable. The observers could contribute information about the effectiveness of activities, networking developing during the workshop, the effectiveness of the leaders, and other unforeseen benefits and drawbacks.

Identifying Data Sources

After identifying common themes in the questions generated above, the evaluator places draft questions into a matrix against potential data sources, as in Figure 3, to determine which sources are likely to provide information for each question. This process identifies many potential sources for responses to each question. For example, to investigate whether participation in the seminar fostered cross-institutional cooperation, we could gather data from the state board, all the institutions, and the participating faculty. If we expand our definition of cooperation to include networking, some data might be gathered during the seminar itself by observing instances of networking. Such observations could come from staff or from non-participant observers as well as from participants themselves. On the question of cost, information could be gathered from the

State Board, the host institution, the staff leaders, the invited institutions, the participants, and relevant documents. Whether or not the goals were met could be assessed from the standpoint of the State Board, the faculty, the staff leaders, the invited institutions, the participants, and documents.

Deciding When to Collect Data

By crossing sources with times for each question, as was done in Figure 4, we consider which data might be collected at each of several points in the program. Prior to the start of the program, we would get information about program goals, baseline data on networking and inter-institutional cooperation, entry level assessments of participants' content knowledge and teaching expertise, participants' expectations and goals, and so on. During the program itself, we could monitor gains in content knowledge and teaching skills by getting reports from host faculty and staff and from non-participant observers. The latter could also gather information about networks that develop within the group during the program. Immediately after the seminar, we could gather data from the leaders on their reactions to the program and on networking that developed, from the participants on their self-evaluation of goal achievement, and from observers on the overall program. After a delay, the State Board, the host and invited institutions, and the participants could provide information on continued networking; documents proposing new networking activity would also be relevant data here. At this point, colleagues and students would have had time to form impressions of improved teaching or content knowledge as well.

Collapsing across questions also gives us an idea of how many measures are being asked of each source. One could not expect the host faculty, for example, to provide extensive data unless they were compensated or had a vested interest in the program.

Choosing Data Collection Methods

Once questions and data-gathering occasions are identified, the evaluator considers ways of gathering the data. The continuum suggests techniques worthy of consideration. At the emic end of the continuum are interviews held at the beginning and end of the program and as delayed follow-up. These data are checked against application forms and action plans. Observer narratives during the seminar can track development of networks among participants, and semi-structured group maps filled out by participants can indicate how they perceive relationships that

developed in the group. Later, more structured questionnaires can provide follow-up data about contacts outside the seminar. Methods from different points on the continuum can focus on the same question.

Etic methods tend to be more familiar and readily accepted by participants and funding sources, and are generally less labor-intensive for participants and evaluators. Emic methods often reveal causal relations and aspects of the situation that are not apparent from more structured data-gathering techniques. Thus, the two types of methods complement each other.

Putting It All Together

Ultimately, the evaluator settles on measures that are feasible and affordable and that do not unduly tax respondents. In Figures 6-8, we show some of the resulting matrices. The slice in Figure 6 represents data collected for a given question. Slicing the cube in this way by question allows the evaluator to see how each question would be pursued. Slicing according to source (Figure 7) helps to organize data collected from each group. For example, data on several questions collected in questionnaire format from a single source could be combined into one questionnaire, so that the source is contacted only once per period (once at the beginning, once at the end, and so on). Slicing according to time (Figure 8) facilitates the evaluator's time management, so that procedures can be arranged to collect the appropriate data at a specified point in the program.

By consulting this matrix during both planning and execution phases, evaluators can cross-check data, thereby increasing the likelihood of obtaining multiple sets of information for each question and from each source. We think this makes for a more efficient and thorough evaluation process.

Advantages of the Circular Model of Evaluation

The circular model forces those who plan and evaluate programs into a creative and open stance. With this model, evaluators who tend to be naturalistic become more systematic in exploring potential data sources, and those tending to be rationalists become more expansive in exploring alternative data sources. By opening up new ways of thinking about evaluation and by taking a broad approach to evaluation, the circular model avoids the extremes that sometimes characterize linear approaches. These extremes require choosing either only those questions for which one has data or collecting only those data for which one has

questions. One advantage of the circular model is that it avoids stagnation and encourages creative exploration.

Another advantage is that the model reveals multiple perspectives. As the example shows, the model utilizes multiple sources and forms of

Question: Does the program affect teaching?

Source	Occasion			
	Before	During	Immediately after	After a delay
State Board				
Host Institution				
Faculty				
Staff		Narrative of Microteach		
Invited Institution				
Participants	Questionnaire or self-report	Analysis of microteach	Questionnaire or self-report	Questionnaire or self-report
Colleagues	Questionnaire or class visit			Interview or class visit
Students	Course Survey			Course Survey
Observers	Critique of pretape	Analysis of microteach	Critique of final tape	
Documents	Applications Personnel file	Microteach videotapes

FIGURE 6. Evaluation Slice: Time by Source for Question "D"

data. One data collection technique may reveal trends not apparent under other techniques. One data source may be inconclusive while others are less ambiguous. This model ameliorates the common naturalistic problem

Source: State Board

Occasion

Question	Before	During	Immediately after	After a delay
A. How much does it cost?	Estimates	Cost sheets	Summary of costs	
B. Does it increase content knowledge?				
C. Does it increase networking?	Counts of cross-school projects			Proposals for cross-school projects
D. Does it increase teaching effectiveness?				
E. What is the quality of the program itself?			Letters from participants	
F. Were the goals met?	Transcripts of initial meetings		Summaries	
G. How much staffing is required?		Time sheets

FIGURE 7. Evaluation Slice: Question by Occasion for Source "State Board"

of unreliability of a single method or source. It also avoids the rationalist’s inability to exploit serendipitous findings when those findings are not accommodated by previously specified questions. Multiple perspectives permit data triangulation and lead to more accurate conclusions.

Occasion: During the program

		Question							
Source		A	B	C	D	E	F	G	H
State Board		costs							
Host Institution		accounts							support log
Faculty			questions	logs	microteach	reaction sheets			time log
Staff			questions	logs	microteach	complaints			time log
Invited Institution		travel							
Participants		expenses	logs	logs	self-report microteach	midway evaluation			
Colleagues									
Students									
Observers			questions	observe logs	microteach	observe			time log
Documents			thought papers	joint projects	videotapes	midway questions

FIGURE 8. Evaluation Slice: Source by Question for Occasion “During the Program”

A third advantage of the circular model is its more accurate depiction of the evaluator's experiential situation. Evaluators cannot escape pre-set questions, even while attempting to be open to the natural setting. Nor can they be free from pre-existing knowledge and expectations that influence the selection of questions. A window of creative opportunity opens when they recognize the value of both naturalist and rationalist approaches.

Finally, the circular form of the model recognizes that evaluation is a never-ending process. Within each phase, activities influence one another. For the model as a whole, one cycle may yield new questions or identify the need for new data, and thus initiate another cycle. One can imagine cycle upon cycle moving toward presumably more accurate and usable findings. In practice, evaluation ceases when results are sufficient or when circumstances force a stopping point; but in theory, evaluation never reaches its end.

Conclusion

The circular model of program evaluation assists evaluators to be more creative and yet more efficient in their evaluation projects. By using tools of creative design, the evaluator is stimulated to expand the repertoire of possible data collection techniques. The resulting evaluation is realistic, but still uses information that would be considered appropriate under both naturalist and rationalist paradigms.

References

- Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology*, 29, 75-91.
- Tyler, R. W. (1950). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.