DOCUMENT RESUME

ED 224 813 TM 820 815

AUTHOR Thompson, Donald E.
TITLE School-Based Evaluation: A Stakeholder's Approach.
PUB DATE Mar 82
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Audiences; Evaluation Criteria; *Evaluation Methods; Field Studies; Formative Evaluation; *Models; *Program Evaluation; Program Improvement; Research Utilization
IDENTIFIERS Evaluation Utilization

ABSTRACT

To examine programs currently in place, an evaluation model was designed to insure inclusion of all those who have a stake in program performance. The evaluation design includes: (1) a set of researchable questions which are to be answered by the evaluation, each question referenced to one or more appropriate audiences; (2) for each question, the items, measures, and data sources to be used, with empirical estimates of quality for each item-source combination; (3) the collection procedure (instruments and user guides) to be employed for each item, and a schedule for collection; (4) a sampling plan for all samples to be used in the evaluation; (5) an analytical plan, to include data maintenance and quality control, aggregation rules (for items, constructs, program components), and statistical treatment; (6) a reporting plan, tailored to the needs of each audience; and (7) a complete management and staffing plan to implement the evaluation design. The implementation of this evaluation model is discussed in the framework of field studies of programs in place, an evaluation of the needs of users, technical limitations on the design, and the formulation of the design. (Author/PN)

SCHOOL-BASED EVALUATION: A STAKEHOLDER'S APPROACH

Donald E. Thompson, Ph.D.

Director, Division of Health, Education,
and
Social Professions

The University of Michigan-Flint
Flint, Michigan

Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York. March, 1982

Donald E. Thompson, Ph.D.
Director, Division of Health,
Education, and Social Professions
The University of Michigan-Flint
Flint, Michigan

SCHOOL-BASED EVALUATION: A STAKEHOLDER'S APPROACH

In a time of dwindling resources, the external evaluator is called upon more and more by superintendents of schools, school boards and school district evaluation offices to conduct program evaluation. While external evaluators do bring to the assessment process "objectivity" (often too cold) what they very often are unable to bring is "subjectivity" necessary for interpreting data in the context from which it is taken and for which is will be used. Often the school system is given a two dimensional evaluation model which looks at the impact of the program on the provider and providee in some kind of pre-test/post-test mode -- a purely summative model which often lacks discriptiveness and provides no information for all the "users" or those who have a stake in the program evaluated.

The following model is a stakeholder model used by this evaluator for examining "programs currently in place." This model attempts to insure the inclusion of all of those who have a stake in program performance.

The design of the evaluation model should include:

- the set of researchable questions which are to be answered by the evaluation, each question referenced to one or more appropriate audiences,
- for each question, the items, measures, and data sources to be used, with empirical estimates of quality for each item-source combination,
- the collection procedure (instruments and user guides) to be employed for each item, and a schedule for collection,
- a sampling plan for all samples to be used in the evaluation,
- an analytical plan, to include data maintenance and quality control, aggregation rules (for items, constructs, program components), and statistical treatment,
- a reporting plan, tailored to the needs of each audience, and
- a complete management and staffing plan to implement the evaluation design.

Task 1: Field Study of the Program in Place

In order to accomplish the task of examining a program in place, there are three essential tasks. The first and overriding requirement is to assemble as much descriptive information as possible about the operating program. The second is to obtain selected information about the context in which the program operates. The third is to arrange this descriptive information (program and context) in the form of a program rationale. Each of these sub-tasks will be discussed in turn.

1.1 *INVENTORY OF THE PROGRAM INPUTS*

This is the most deceptively simple task in program evaluation. No design or analytic chores are involved, just a catalogue of what has happened and what is planned. It looks easy. It is not easy at all; many evaluations flounder at this point, though the failure is seldom recognized. Unless the basic program description is complete, completely accurate, and available for inspection by all parties, there can be no agreement on *what it is that is being evaluated.* Failure to reach this agreement early guarantees that the latter results will be challenged on the grounds of relevance.

The implications for the conduct of the work are:

(1) *Disaggregate the program descriptions into as many discrete components as possible.*

The melange of activities which comprise most programs will be found to possess common elements. But before deciding on those themes, the factual description should concentrate on the smallest feasible unit--probably, judging from what we know now, each specific activity within the program. The product will be a large loose-leaf compilation, easy to update and expand by bits.

(2) *Take measures to ensure that future plans and current actualities are carefully distinguished.*

In a fast-moving program, there is an understandable tendency on the part of program staff to discuss an activity in terms of what the respondent "knows" it will be like in a few more weeks. The demarcation between present and future becomes blurred. The evaluator must force the distinction and must develop instruments which are sufficiently flexible to deal with both present and future (planned) activities.

(3) *Rely on interviews and observation rather than on archival data.*

In general, I have a strong predisposition to use archival data. But many school programs suggest an exception. Recent experience with complex and dynamic programs suggest that the written descriptions and records tend to be incomplete, omitting important activities that were never written down, and even unintentionally misleading, giving figures (e.g., "35 students are enrolled in...") that were *expected* to remain true at the start of the semester, but that turned out to be wrong. By their very nature, many social action school-based programs are concerned with doing, not with documenting.

Recognizing the intrinsic difficulty of the descriptive task, and realizing that much of the essential information will be located only in the heads of the program developers and managers, it is necessary to depend heavily on direct interaction with program staff and extensive observations of the program in operation. It is also necessary to explore current operations, planned expansions and modifications, and get some of the contextual flavor of the program.

1.2 *DESCRIPTION OF THE DEVELOPMENT AND CONTEXT OF THE PROGRAM*

The second of the subtasks is a description of how the program got to be where it is, the allies and adversaries it picked up along the way, and the context within which the program is lodged.

The question the program staff must answer early in the process is: how much does it really want to know about these topics, and why? The genesis and development of any major program presents many opportunities for studies which are "interesting" in an academic sense, but which provide no really useful information to decision-makers. These "opportunities" must be declined.

The high-priority topics should include the following. History

bearing on *acceptance of an assistance to the program* by funding sources, the school system, and local community organization involvement should be developed in detail. All literature relating to *analogous programs* for students should be assembled, including curriculum/instructional innovations and service/counseling oriented approaches. Local history should be detailed, and *theoretical literature* should be surveyed. The evaluation staff itself must possess considerable expertise on these topics at the outset.

The evaluation must be on the alert for only those specific elements that help to illuminate the program's goals or performance. Avoiding the trivial and focusing only on the items of crucial relevance, is an art, not a science. The only test is consensual; reasonable people agree that a particular element is important while another is not. Like the program descriptions, the presentation of contextual variables should also be public--open to inspection by all interested parties.

1.3 *DEVELOPMENT OF THE PROGRAM RATIONALE*

The final subtask of Task 1 is to array the facts on the state-of-the-program into a framework for subsequent development of the evaluation design.

For a simple program, a rationale can be constructed that fits quite closely with the model described earlier: a fully articulated map of outcomes that stretches from initial program inputs to ultimate outcomes. Drafts of the rationale can be circulated to program staff and to stakeholders in general, and revisions mutually agreed upon in an alternative process.*

---

*"Stakeholders" is a convenient term coined by Guttentag and Edwards (1975) to denote persons who have a stake in the program and its evaluation.

Often this neat, self-contained process is not feasible when the program is complex. There is no unitary map of inputs, immediate outcomes, intermediate outcomes, and ultimate impact. Rather, there are as first products, agreed upon *statements of objectives* and *hierarchies of outcomes*, serving as a framework for a next set of decisions about what to evaluate and in how much detail.

The first priority in this process is to establish that the framework is *accepted as a fair and complete statement by the program staff*. As an evaluator, this is the first major objective. The goal is to add something to program staff's understanding of themselves. The ideal state of affairs has been reached when the program staff decides that the evaluator has put on paper a formulation of the program that is better and more complete and insightful, that the program staff has had time to do for itself.

1.3.2 *HIERARCHIES OF OUTCOMES*

The activities just described will focus the evaluation on the components and the objectives that are central to the program, and hence to the evaluation. In terms of the program rationale, they help define and delimit the far left- and right-hand elements of the program evaluation model. They also provide the skeleton for the intervening elements.

Task 2: Evaluation Needs of Users

The importance of talking to users early in the evaluation

process has always been recognized as an approach to evaluation. Typically, however, we tended to define "users" narrowly, to include the program people and the sponsoring agency. Further, the interactions tended to be unstructured: the evaluation project director would stop by offices on field trips to let the local users know what, in general, was happening. Last year, my consulting group had occasion to undertake an evaluation that required more formal interactions with a variety of user populations, at both the national and local levels. The experience was often frustrating, always instructive, and ultimately rewarding. We believe that we have learned in the process. What we have learned we feel should be incorporated into the approach when evaluating school programs; some false starts can be avoided.

General Observations on the Objectives

The first point is a short one: the central concept of structured interaction with stakeholders is sound. Informal, *ad hoc* meetings have their place, but the process must also incorporate systematic procedures for ensuring that prospective users of the evaluation have had their say about what is needed and when.

But, more specifically, what are the realistic objectives of these interactions? They depend on who the users are: *program staff, institutional patrons* (existing or prospective), or citizen *participants/consumers* in the programs. We discuss the objectives for each in turn.

Program Staff. In the first stage of interaction with the staff, the objective is integral to the general design objectives: to find out from program staff *what they see as legitimate, comprehensive measures of success and failure,* without regard to whether they are quantifiable or otherwise measurable. It is important, however,

also to inquire of program staff how *they* judge their progress or lack of it, when they do not have an evaluator around. Frequently, the observational indicators that program's staff use informally, almost unconsciously, can be translated into systematic measures.

There are many other, related comments to be made about the evaluator/program relationship. We reserve them for the discussion of *Other Factors*.

Institutional Patrons, Existing and Prospective. On first examination, it seems very simple: meet with the people who eventually will be making policy decisions about the program, and advance these statements:

1. Evaluations typically are not used. You know it, and we know it. Let's try to change that situation.
2. What are the decisions you will eventually have to make about your educational program?
3. What do you want to know to make those decisions?
4. When do you need to know it?
5. In what form would the information be most useful to you? A final report? Periodic interim reports? Briefings on specific topics, on demand?

And those five questions happen to form the agenda.

In short, we judge that the process increased the likelihood that the evaluation will be read attentively by some of these key persons, and that is not a trivial virtue. But more can be gained.

The key to improving the benefits of these meetings is *to give the institutional stakeholders something to react to,* rather than asking them to fill in a blank instruction sheet for the evaluation. For example, do not give them a long roster of potential outcome measures (as we did). Rather, wait until the issues to be given

priority are starting to crystallize, then propose a draft design to the group. Do not ask for suggestions about the most useful format for results; devise some specific options, describe them in some detail (perhaps even with mockups) and ask for responses. And finally, when trying to ascertain the group's priorities, take a reduced set of dimensions and obtain specific ratings and rankings of their relative importance (e.g., via the type of procedure described by Guttentag & Edwards, 1975).

Participant/Consumers. Many school programs are structured in a way that it should be easy to assemble participant/consumer panels for discussing evaluation issues with project staff. The steering committee is the natural focal point for these discussions. The objective includes both information-gathering (what do the parents, students, and community persons see as the crucial measures of success or failure; when do they realistically expect occurrence of the various levels of outcome) and some degree of information dissemination on the evaluator's part. Even more than policy-makers, participants in the program have either been left completely out of the evaluation's audience, or have come to perceive evaluations as a statistical flimflam with very little of substance to say about a program that they see at first had on a day-to-day basis.

Procedures

Interactions with users will occur in three ways during this process.

1. *Intensive, semi-structured interactions with program staff and participants/consumers about the program.* This coincides with the early

stages of Task 1. The evaluators should forego even preliminary attempts to determine user needs until they have a firm grip on what the program is about and how it is operating at the demonstration sites.

2. *Structured interviews with key institutional patrons and program staff about evaluation alternatives.* Asking the "five questions" of a group proved to be unproductive. We believe useful results *can* be obtained when they are raised in a one-on-one interview situation. Considerably greater candor about the realities of institutional and political constraints should be forthcoming.

3. *Presentation of a draft design.* A draft design with the kinds of specific proposals discussed earlier will be presented separately to each of the three groups, meeting as groups.

Subsequent interactions would include presentation (by meeting or mail) of the final design and periodic updates on progress and issues. The appropriate nature and extent of these subsequent interactions is, of course, one of the issues to be decided by the initial ones.

Task 3: Technical Limitations on Design

The objective of this task is to specify the nature of the evaluations which are (a) potentially doable in context, (b) not possible in that context, and (c) both doable and useful in that context. The third category considers both methodological adequacy and the requirements of the several audiences for the for the evaluation.

Having described the program in great detail (the program rationale) and having determined the needs for evaluative information throughout the system, the design task proceeds to compare

the two. For each user requirement, we know (a) in which "segment" the program rationale the relevant activities or events occur, (b) the hypotheses which link those events to antecendents and consequences, (c) the process variables and disposing conditions which mediate the relationships, and (d) the relevant data which are already generated by the system.

For each hypothesis at issue, we ask:

- are any external comparisons feasible; do adequate comparison exist, and if so, can they be accessed,
- what are the alternative hypotheses which might be invoked to account for an observed X_1-X_2 relationship,
- can multiple indicators be identified which would support a convergent validity argument,
- since the hypotheses are embedded in a larger causal network, can nearby portions be tested by more power means,
- how have other evaluations dealt with instances of this type, and finally,
- all things considered, what is the best test (or set of tests which can be applied?

The answer (for each element of the evaluation) can be assessed by reference to:

the available literature on interactive evaluations,

outside experts in evaluation.

The outside experts will assess the adequacy of the reasoning which is used to reach conclusions and will suggest modifications should the logic appear weak. Given that the logic is adequate, the reviewer's task is to propose a more powerful test than the one advanced by the project staff. The process becomes a dialogue between the designers and the reviewers.

The product of the task is a set of evaluation activities

keyed to user requirements and documented as to why each is recommended as the best available solution.

Task 4: Formulation of Design

The formulation of questions to be answered by the evaluation begin as soon as the rationale has started to take shape. As more information accumulates about the program, about key stakeholders, about the three environments, and about evaluation needs, the questions are continually sharpened. The process of establishing priorities also should begin early and continued throughout. The final iteration will therefore not be a major task; the final list of questions will be available for review by users later in the project. These questions will be used to organize a preliminary data handbook.

The data handbook will also include a crude flowchart to indicate what items of information are to be delivered, where and to whom. This information, together with the recommended indicator-source information, will lead directly into the development of instruments. While one cannot know how many separate instruments will be required, one can be fairly certain that needed will be some number of interview guides, forms for retrieving data from archives, observational checklists, incident report forms, and possibly questionnaires. For each instrument, a complete user's manual need be developed. Quality control procedures for each step of the collecting-recording-processing-reporting sequence need also be established.

A program that is worth a major evaluation typically stems from a few central concepts. The program tries to operationalize to make them work. The long-term scientific value of conducting

an evaluation, in my view, is to learn something about the validity of those concepts and to make suggestions that try to bring practice more closely in accord with the expressed concepts of the project. In using a model that takes into account the participants, users and patrons of the program -- the above can more rationally be accomplished.

But, second, we think much of the current theoretical debate about propriety on the detachment/involvement issue is irrelevant. From a practical standpoint, a hands-off, detached clinical stance is usually out of the question, certainly so in the program evaluation. The evaluators are going to be deeply involved with the program staff, or they will be cut off from the kinds of data and kind of understanding necessary to carry out a meaningful relationship.