PROGRAM EVALUATION AND PROGRAM MANAGEMENT*

Harry S. Havens, *U.S. General Accounting Office*

Ideally, there should be some useful relationship between the process of managing a program and the process of evaluating it. We rarely find that ideal in the real world. The purpose here is to outline what managers and evaluators must do if they are to work together and to suggest some reasons why it is essential that they do so.

It may be useful to consider, first, what is meant by the term "program evaluation." It is a much abused label and, unfortunately, conveys many different things to many different people. In addition, because of a relatively brief but checkered history that includes a substantial number of bad program evaluations, it carries a lot of excess baggage.

Because of the fuzziness which seems unavoidable in any effort to "define" program evaluation, it seems better to approach the task by describing what it does—or at least aims to do. For purposes of this discussion, let us agree that a program-is a collection of activities intended to achieve a common purpose. The process of program evaluation, then, is an effort to judge the extent and efficiency of accomplishment and to find ways of improving it.

A "good" program evaluation, like a "good" program, is one which accomplishes its purposes with reasonable efficiency. The common purpose sought by any program involves making some change in the real world. That is, the intended results are external to the program. The same is true of program evaluation. An evaluation may meet all the standards of rigorous design, careful data collection and analysis, and a beautifully written report. If it does not affect the real world, if it is not used, it has failed the test which evaluators themselves apply to the programs they evaluate.

But the real world which the evaluator usually seeks to affect is the program itself. He does so by affecting the decisions which are being made about that program. It is this central purpose of most program evaluation activity which necessitates its linkage to program management. Generally speaking, program evaluation serves little purpose if it exists in a world unto itself, isolated from the process of program management.

Those processes go well beyond the individual who may be identified as the program manager. To be realistic, the concept of program management must embrace all those decisions and actions which impinge on the program, from whatever source. The Congress is engaged in program management when it enacts, amends, or repeals laws governing the program or governing the actions of people who administer or participate in the program. The president and the Office of Management and Budget (OMB) engage in program management when they recommend legislation and funding levels for a program, or promulgate rules which affect it. State and local governments engage in program management when they exercise their discretion to decide whether and how the program will function in their jurisdictions.

Given this broad concept of program management, where should the linkage with program evaluation occur? Should the evaluator seek to affect the real world of the Congress? the Executive Office of the President? the agency head? state and local government? or the person charged with administering the program? The answer is any or all of the above, depending on the issue or issues being addressed. The evaluator should seek to have the results of his work used by whoever is in the position of making a decision to which the evaluation is relevant.

If the efficiency of internal operating procedures is at issue, the evaluator must connect with the program administrator. If the adequacy of a law governing the program is at issue, the evaluator must face the fact of a multiplicity of decision makers, including the agency head, OMB, the president, and the Congress. Each of these sets of potential users has needs which differ. The evaluator who wishes his work to affect the real world of the program must be attentive to those differing needs. If those needs are in conflict, and they may well be, the evaluator must seek ways of reconciling them. Failing that, the evaluator must reach a

---

*This article was adapted from the Roger W. Jones Lecture, delivered by the author at American University on February 27, 1981.

Harry S. Havens has been assistant comptroller general for program evaluation of the U.S. General Accounting Office since April 1980. He was director of GAO's program analysis division in various positions in the Bureau of the Budget and Office of Management and Budget.

judgment as to which needs are most important and design the evaluation to satisfy those needs.

To say the least, this need to identify the intended audience and to design the evaluation around the needs of that audience makes life rather difficult for the evaluator. It has been known to cause a mild form of schizophrenia in practitioners. Life can be even more difficult, however, if the evaluator is one who does the work first and only afterward (if ever) thinks about the intended audience. That evaluator can look forward to a career which is likely to be short and almost certain to be full of frustration.

The evaluator who takes the problem of utilization seriously, however, may be tempted to throw up his hands at the apparent impossibility of trying to identify the intended audience before he knows enough about the program to judge, even tentatively, what may be wrong (or right) with it. But the problem really is not that difficult. Rarely, if ever, does an evaluator set forth with a blank slate, attempting a "complete" evaluation of a program. (An evaluator with that concept of his role should be given a stern lecture about the evil of hubris and then required to write "pride goeth before a fall" one thousand times before leaving the room.)

When an evaluation turns out to have been useful, it can usually be traced to the fact that it succeeded in answering a specific, clearly defined question, a question someone wanted answered. Therefore, the evaluation process should start with an attempt to articulate such a question. One hopes there is someone interested in the answer to that question, and it is usually possible to find out who and why. (If it turns out that no one is interested, the evaluator can save himself and everyone else a good deal of time, energy, money, and paper by starting over again with another question or another program.)

In many cases, perhaps most, the evaluation activity is stimulated by evidence that someone is interested in the evaluation of a program, or some aspect of it. The evaluator then translates this expression of interest into the evaluative question. If the expression of interest was properly understood, one can expect the client to be interested in the answer to the question. One should double check, however, to avoid the subsequent unpleasantness attendant on having misunderstood the request or other indication of interest.

Once the interested party or parties has been found, it is important to find out why they are interested, that is, what they expect to do with the answer. If the expectation has nothing to do with making a decision, one should be rather pessimistic about the likely utility of the report. A lot of evaluation effort is wasted answering questions which have no bearing on decisions, questions asked out of idle curiosity or a desire to keep the evaluators out of mischievous activities. One hopes the results are a useful contribution to basic research, but there is not much evidence of this, either.

In some cases, however, the person who wants the answer will want it for a very practical reason. The evaluator should seize these opportunities with great enthusiasm, for they tend to be rare. But, notwithstanding his joy at finding a candidate for useful evaluation, the evaluator is

well advised to probe a bit deeper. If the question relates to a specific decision, he should find out who will be making that decision. It may well not be the person seeking the answer to the evaluation question. Rather, that person will be planning to use the evaluation results to influence a third party (or parties) who will actually make the decision. The head of an agency, for example, may want the evaluation as the basis for legislative proposals which will ultimately be considered by the Congress. In this case, the evaluation must be planned around the needs of the third party, not just the needs of the requestor.

It is also essential to find out, if possible, when the decision will be made. If the evaluation results cannot be delivered in time to be used, there may be little point in producing them at all. If time is a problem, however, the evaluator is obligated to look for ways of solving it. For example, preparing a formal written report is often a time consuming activity. The evaluator may be able to save this time by presenting the results orally. Even if these results must be characterized as tentative, they are likely to be better than nothing at all.

. . . program evaluation serves little purpose if it exists in a world unto itself, isolated from the process of program management.

The evaluator must also be sure that the question is answerable or find some way of refocusing it in a way which is answerable—and still useful. Answerability has several dimensions, and the evaluator must be conscious of all of them. There are some questions, important ones, which we simply do not yet know how to answer. Others we can answer only in rather imprecise terms, and the answers are about as helpful as they are precise. In-other cases, we know how to answer the question, but the precision of the answer, and our confidence in it, is a function of the time and resources available. There may be a fourth category, one in which reliable, precise answers can be obtained both quickly and cheaply. If this category exists, however, it is rarely encountered and probably involves answering some rather unimportant and uninteresting questions.

When the evaluator faces an important but unanswerable question, his responsibility is rather straightforward. His first obligation is to be honest with the client. He must explain the problem to the requestor and seek agreement on some other question (or some derivative of the question) which is both important and answerable. One hopes the requestor will accept the situation with good humor, but that is not always the case. (The world is still populated by those who would prefer to behead the messenger rather than accept the bad news, a fact to which any experienced evaluator can readily attest.)

The case in which answerability is a function of time and resources can become even more difficult to handle. It requires the evaluator to enter into an often complex process of negotiation with his client. The evaluator has a professional responsibility to assure that the client understands the limits on answerability imposed by constraints on time and resources, so that the client will have reasonable expectations about the results of the evaluation. At the same

time, however, the evaluator must avoid being so negative and purist as to cause the client to lose interest in what may be a very useful project.

An evaluation which is less than perfect because of limited time and resources can still yield useful results. The utility of the findings, however, is directly related to the ability of the evaluator to provide information (however qualified it must be) which is relevant to the decision which must be made. Thus, the evaluator must walk a very narrow line. He must seek to be as helpful as possible to the client without compromising his professional responsibilities.

The process of identifying a potential user, and then defining a question which is both relevant and answerable within the limits of available time and resources can be particularly difficult for an independent evaluative organization such as the General Accounting Office (GAO). With respect to about two-thirds of its work, the decision on what to review, and when, is made through GAO's internal planning system and is guided by its basic legislative requirements. This independence is clearly a vital asset. But it carries with it a risk. The matters which GAO considers relevant may or may not be seen in the same light by its primary client—the Congress.

In order to minimize this risk, GAO engages in extensive dialogue with key committees. This serves several purposes. First, it permits the adjustment of plans in recognition of congressional needs and schedules without impinging on GAO's statutory independence. Second, it provides an opportunity to gauge the likelihood that the work will be used and thus to judge whether or not the level of investment is warranted. Finally, the discussion sometimes influences the committee agenda, leading to the consideration of issues which might otherwise have been overlooked.

A lot of evaluation effort is wasted answering questions which have no bearing on decisions, questions asked out of idle curiosity or a desire to keep the evaluators out of mischievous activities.

This might be an appropriate point at which to mention the subject of "lost causes." There are times when GAO undertakes a review knowing full well that there is little likelihood of the recommendations being implemented in the short run. This activity is not born of a masochistic desire to be unpopular or a failure to recognize the importance of relevance. Rather, it comes from a conviction that, in time, the cumulative weight of evidence can change the boundaries of political feasibility.

When GAO undertook its review of the Davis-Bacon Act, the prospect of repeal or substantial change was remote, to say the least. Today, it is a little less remote. When GAO first recommended that Treasury collect interest on money in commercial bank tax and loan accounts, the idea was rejected. It has since been accepted.

Most of the time, however, evaluators cannot afford to define relevance in this extended fashion. They must earn their keep by being useful to decision makers today. This is the basis for insisting that the first task of the evaluator is to define a question to which an answer, useful to an identified client, can be produced within available time and resources.

Once the evaluator has done this, he can proceed to do the work for which he was presumably trained. He can start trying to answer the question. This will not be easy, either, but at least he has been trained (one hopes) to solve the problems in this part of the job. He can sally forth in search of data which he can subject to various obscure forms of analysis which, in turn, will permit him to write a report which may be of immense interest to other evaluators and, all too often, to almost no one else. He may do this very well, for it is what he was trained to do.

Having done so, however, the evaluator who is still committed to effecting change in the real world faces the task of reentering that world. That task is difficult, even for those who have done the first part well. One hopes, for example, that the issues have not been overtaken by events, that the requestor is still interested, has not been replaced by someone else, still remembers the terms of agreement under which the evaluation was undertaken, and still considers the evaluation results relevant to the decision which must be made. The reentry process is more likely to be successful if the evaluator has maintained contact with his client, providing interim results and making interim adjustments to the design which are as responsive as possible to the client's evolving needs.

This effort to assure continued relevance (and to remind the client that the evaluator has not retired) serves another purpose as well. It is likely to have given the evaluator some practice at translating his results into words which someone other than an evaluator can understand. This is one of the most difficult parts of the reentry process. Communicating effectively the results of an evaluation can be just as fraught with problems as deciding what to evaluate and how. Evaluators have only begun to understand these problems and are nowhere near solving them.

Recently, there has been greater emphasis on improving the quality of written products. This has focused on such matters as improved clarity in writing (avoidance of technical jargon, etc.) and greater use of abbreviated summaries. But the focus on written products is, itself, part of the problem. The focus, instead, should be on the process of communicating, in which written reports play an important, but by no means exclusive role. Of equal—perhaps greater—importance is the evaluator's ability to convey information orally, and to do so clearly and concisely. One often encounters decision makers with whom it would be futile to attempt to communicate in writing. Some simply do not like to read or, because of confidence in their ability to judge people in a face-to-face setting, may prefer to receive information orally. For others, preference has little to do with the matter; they would not have time to read if they wanted to.

The evaluator who wants his work used must adapt to the operating style of the decision maker. If the decision maker has no time to read, there is little point in sending him a written report. If he has five minutes of reading time, send him five minutes of reading material. If he only has time for

five minutes of conversation while going from one meeting to the next, use those five minutes wisely.

This does not mean that detailed, extensive written reports are unimportant, or that the evaluator can dispense with them casually. Rather, it means that they are rarely the most effective means of communicating with the decision maker. If well-prepared, a formal written report can still serve other important purposes. It permits communication with a broader audience, those concerned about the program, who may help shape attitudes about it and influence its direction over a long period of time.

A report also facilitates communication with professional peers, whose suggestions and criticisms will help one do better work in the future, and whose opinions largely determine one's individual and institutional credibility. Finally, a written report serves as a record of what was done. This makes much easier the process of judging validity and, if used properly, permits other evaluators to avoid reinventing the wheel. Useful as these functions are, however, none rivals in importance the need to find the most effective means of telling the decision maker what he needs to know, when he needs to know it.

The discussion so far has addressed the responsibilities of the evaluator toward the manager. But the manager has responsibilities too. If the process of evaluation consistently fails to yield program improvement, the tendency is to assume that the evaluator is at fault. No doubt this is sometimes the case. But it is equally clear that the fault may well lie elsewhere—with managers who do not make use of evaluations. Some do not want to do so; others just do not know how.

Managers are rarely trained in the technical aspects of evaluation and it is pointless to wish they were. Indeed, it is not at all clear that a good evaluator would necessarily make a good manager or vice versa. The skills are quite different and it is unusual to find them embodied in the same person. Incidentally, this says something about the need to be attentive to the difficulties of managing an organization whose mission is the performance of evaluations. In this case, the managers are usually drawn from the ranks of evaluators.

But, while most program managers cannot be expected to be technically expert in evaluation, they should have a basic understanding of the subject. After all, managers are rarely personnel specialists, but they must understand the system; they are rarely budgeteers, but must understand a budget; they are rarely accountants, but must use accounting data. As an ingredient in the makeup of a good manager, the ability to understand and use evaluative information is just as important as any of these other skills.

This does not require a high level of technical understanding. The manager does not need to be able to perform a regression analysis or a chi-square test, any more than he need memorize all the rules governing selection and promotion of personnel. But he does need to understand that there are techniques for the systematic analysis of quantitative data, just as he needs to know that there are rules involved in a merit personnel system.

More important than any technical understanding, however, is a conceptual understanding of evaluation as a re-

search process applied to the answering of questions. With this conceptual understanding must come a recognition that the manager shares with the evaluator the responsibility for defining evaluation questions which are relevant to the manager and answerable by the evaluator. The matter of relevance, particularly, is one for which the manager should assume a very heavy measure of responsibility. Only the manager can know what questions are relevant to the decisions he must make. Frequently, even he cannot be sure, but he is in a much better position to judge the relevance of a question than is the evaluator.

Left to his own devices, the evaluator can only speculate on the matter of relevance. Yet all too many managers, all too frequently, leave the evaluator in the dark. Later, they demean the process of evaluation for its lack of relevance to real world problems, ignoring the fact that it was their own behavior which led to the examination of irrelevant questions.

The matters which GAO considers relevant may or may not be seen in the same light by its primary client—the Congress.

Along with the responsibility to take part in defining a relevant and answerable question goes the responsibility to listen to the answer. Listening, of course, is an art in itself, one in which managers should be proficient. It does not mean having the time or inclination to plow through a 500-page report alternating between turgid prose and technical jargon. Even less does it mean uncritical acceptance of findings, conclusions, and recommendations.

In this context, listening means the active process of obtaining from the evaluator the key elements of information which the manager needs in order to make a decision. At a minimum, this means learning the answers to the evaluation questions. It also means, however, that the manager must learn enough about how those answers were obtained to judge for himself how much confidence he can place in them.

All of this is a lot of work. Why should managers do it? Is it realistic to expect them to take on an added dimension of responsibility, particularly one which seems likely to complicate further the decision process?

In a context of steadily increasing public skepticism about the worth of government activity, it is arguably unreasonable to expect a manager to assume an additional burden out of altruism or a sense of duty and responsibility. Many will do so, but that motivation should not be necessary in this case. Given the environment in recent years, enlightened self-interest, alone, should be enough to motivate an intelligent manager to take an interest in evaluation. That environment is one of intensifying competition for increasingly scarce public resources, at all levels of government. The manager who survives and prospers in this environment, all other things being equal, will be the one whose program is demonstrably effective.

Note both words in that characterization. In the political process which controls the competition for scarce resources, it is not enough that a program be effective, it must be demonstrably effective. Evaluation is no panacea,

but it can be an increasingly useful tool, both in raising a program's level of effectiveness and in convincing others of that level of effectiveness.

The manager who makes effective use of high quality evaluation work will compete more successfully than the one who does not. That conclusion does not rely on an underlying assumption of a super-rational decision process. It only requires the assumption that better information will yield (at least marginally) better decisions. If we do not believe in that degree of rationality, we can dispense with all management processes and make all decisions by rolling dice or flipping coins.

The first dimension of utility lies in the ability to make actual improvements in program effectiveness. On this dimension, the successful manager will be the one who participates actively in defining evaluation questions, the answers to which will permit him to make better informed decisions about how to eliminate barriers to the effective delivery of services. Those barriers may exist anywhere in the spectrum from program design to administrative and operational procedures.

No program is perfect, ever, and a properly focused evaluation will almost always find something which can be improved. The key to success is to view this information as an opportunity to improve, not as a threat. In the long run, the successful manager will be the one who creates those opportunities, through well-focused, internally-generated evaluations, and then makes maximum use of the opportunities when they are handed to him.

. . . the first task of the evaluator is to define a question to which an answer, useful to an identified client, can be produced within available time and resources.

The second dimension of utility involves the role of evaluation in demonstrating effectiveness. This may involve some risk, in that it is a little difficult to demonstrate the effectiveness of a program which is patently ineffective. But this risk has been grossly exaggerated. Few, if any, programs are patently ineffective. One may not like a particular program, believing that the costs exceed the benefits or that the benefits are unwarranted. But that is quite different from saying that a program has zero value. All programs benefit someone. If someone thinks he has found a program without beneficiaries, he should try terminating it. He will soon learn that it represents an essential service to someone in some congressional district.

In trying to demonstrate the effectiveness of a program, the key to success lies in identifying the objectives sought by those who will determine the fate of the program, maximizing that effectiveness (and improvements in it) in terms which are meaningful to those who must be convinced.

This is not as cynical as it may sound. In our system, decisions about the existence and direction of programs are fundamentally—and properly—political in nature. One of the purposes of evaluation is to provide information to be used in that political process. There is nothing wrong with a program manager who does his best to achieve objectives set in the political process and who then attempts to show that he is doing so. Indeed, there is something very wrong with a program manager who behaves differently.

None of this, of course, justifies distorting evaluations in an attempt to demonstrate effectiveness which does not exist. But this sort of cheating has become a much more risky business, anyway. It is difficult to disguise blatant bias, and there is usually someone who has the skill and motivation to detect and publicize the bias.

All things considered, therefore, the manager is well-advised, in his own self-interest, to assure that evaluation is pursued aggressively, to see that it is as balanced and objective as possible, to deal effectively with the problems it brings to his attention and to take pride in the accomplishments it reveals. Doing so will increase the prospects for his survival and that of his program.

Important as these issues are, however, there is a much larger matter at stake than the fate of individual programs. In a very real sense, what is at stake is the ability of government to serve the needs of the people. It is clear that a large part of the public no longer believes in the capacity of public institutions to serve the common good. That loss of credibility feeds on itself. It leads to actions which further impair the capacity of government to act effectively. That, in turn, further reinforces the loss of credibility, and the cycle continues.

We cannot afford for the cycle to continue much longer, but neither is there an easy or painless way of breaking it. One thing seems clear. We in the public service must assume much of the responsibility for the situation and, similarly, we must take on much of the responsibility for fixing it.

For one thing, we have been much too willing to believe in our ability to solve complex social problems and much too reticent to admit that we do not know how, or that it will take much longer and cost much more than anyone has been led to believe. Our own faith in the capacity of government contributed a great deal to the unrealistic elevation of expectations which led inevitably to our present loss of credibility in the eyes of the public. We must balance our confidence in government as an institution with a sense of realism about what government can do well and what it cannot; what it is now doing well, what it can do better, and what it should stop trying to do.

If we are to behave responsibly, it means using every tool at our command—including evaluation—to reestablish this sense of realism about expectations, both in our own minds and in the public. We must be honest with the public. Government can solve some problems, sometimes, but it cannot solve all problems, everywhere, instantaneously. Government is far from useless, but neither is it omnipotent.

We must be open and articulate about the strengths of government as an agent of progress, and about its limitations. Economic problems which have been accumulating for a decade or more can—and must—be solved, but we cannot solve them in one year. Social problems which have faced us for centuries can—and must—be solved, but we cannot solve them in one decade. If we successfully convey these realities about the capacity and limits of government, the public may begin to develop more realistic expectations

of government, neither assuming government can do everything nor, at the other extreme, that it is capable of doing nothing.

As the public begins to adopt more balanced and realistic expectations about the pace at which we can accomplish the properly ambitious goals we have set for our society, we must use every tool at our command—including evaluation —to meet those expectations and to show that we are doing so. Only when the demonstrable effectiveness of our performance begins to match the greater realism we seek in public expectations can we fairly ask the public again to have confidence in us as managers and in government as an institution.