This article raises some questions about the relationship between auditing and evaluation in their approaches to program assessment. How are they similar or different? What can they learn from each other? Four aspects of the relationship are examined: (1) the different origins of auditing and evaluation; (2) the definitions and purposes of both, along with the types of questions each typically seeks to answer; (3) contrasting viewpoints and emphases of auditors and evaluators that emerge in the performance of their work; and finally, (4) some commonalities of interest leading to areas of likely mutual influence and benefit.

COMPARING AND CONTRASTING AUDITING AND EVALUATION
Some Notes on Their Relationship

ELEANOR CHELIMSKY
U.S. General Accounting Office, Washington, DC

With the passage of the Congressional Budget and Impoundment Control Act in 1974, it has been formally recognized at the U.S. General Accounting Office that both auditing and program evaluation are important components of GAO work. Realizing that the products of both professions had become essential for meeting its congressional mandates, GAO pioneered the effort of bringing both together to work under the same roof. An early step toward accomplishing this was taken when GAO expanded its scope of auditing to include “program results audits,” thereby moving toward evaluation activities. A second step involved the creation of the Institute for Program Evaluation in 1980. Currently, GAO is in the process of revitalizing its education and training program to include evaluation design and research methods. It is also making changes in many other areas consistent with recent coexistential experience.

AUTHOR’S NOTE: The views and opinions expressed are my own and should not be construed to be the policy or position of the General Accounting Office.

EVALUATION REVIEW, Vol. 9 No. 4, August 1985 483-503
© 1985 Sage Publications, Inc.
This article, which was presented at the Evaluation Research Society's October 1984 meeting in San Francisco, raises some questions about the relationship between auditing and evaluation in their approaches to program assessment. How are they different? How are they similar? What can they learn from each other? Four aspects that seem important to an understanding of the relationship will be examined here:

1. the different origins of auditing and evaluation;
2. the definitions and purposes of both, along with the types of questions each typically seeks to answer;
3. contrasting viewpoints and emphases of auditors and evaluators that emerge in the performance of their work; and
4. some commonalities of interest leading to areas of likely mutual influence and benefit.

There is a remarkable paucity of published literature on this subject. Consequently, this article relies in large part on my experience at the GAO over the last five years, in seeking to bring auditing and evaluation closer together so that the knowledge of both might be shared and each profession learn from the other.

ORIGINS OF AUDITING AND EVALUATION COMPARED

Given that an audit, in its earliest form, was the rendering of an expert opinion on a financial accounting statement, the origins of auditing lie within the origins of accounting. Members of the accounting profession were responsible for preparing the financial statement under review by the auditors, and the auditors themselves were typically trained in accounting. The origins of accounting, in turn, lie within those of bookkeeping (Hanson, 1970: 131-135); indeed, the terms “accountant” and “bookkeeper” were originally synonymous. Even today certain functions of the accountant continue to be assumed by bookkeepers.

The earliest recorded accountants' association seems to have been the Collegio dei Raxonati, founded in Venice in 1581. By 1742, the city of Milan had established a scale of legal accountants' fees, which certainly
implies some regular use of accountants’ services by that date—at least in Italy.

The beginnings of modern accountancy are usually traced to England, but it is worth pointing out that there exist records for only five accountants practicing professionally in England in 1790, although by 1822 their number had risen to 73. The earliest association of professional accountants in Great Britain—the Society of Accountants—was organized in Edinburgh in 1854. In the United States, higher education for accounting was instituted in 1881 when the University of Pennsylvania’s Wharton School of Finance and Economy was founded; today, schools of accounting exist in every state of the Union. Government regulation of the profession began in New York state in 1896, and the total number of accountants and auditors in the United States moved from about 20,000 in 1900 to some 400,000 at midcentury (Boorstin, 1974).

According to some observers, developments in accounting often have occurred as the result of legislation or of a perceived public need to correct abuses. Boorstin, for example, points out that the accounting profession, as such, developed in Britain “to satisfy the requirements of the New Companies Acts of 1845 to 1868 [and] to protect the public against a company’s false balance sheets.” (Boorstin, 1974: 210). In the same way, “the number of accountants and auditors in the United States remained small until the enactment of a national income tax.” (Boorstin, 1974: 210). The development of authoritative accounting standards had to await the 1930s, although the need for them had been manifest long before that. “Only after the market crash of 1929 did work begin in earnest” (Guy 1981), despite the fact that many of the financial accounting reports and audits of the twenties were notorious for serious flaws—incomplete information, as well as false and misleading information—that more than a few accountants believed to be responsible for many of the unwise business decisions made during the period (Ripley, 1926).

Historically, then, auditing developed as a procedure for detailed checking. Its concern was essentially that of verification, of examining the correspondence between a matter under investigation with a model or standard of operation or performance. (Indeed, the name of the Auditor General of Canada is given in French as the Office of the Vérificateur Général; that is, the general verifier.) Today, although the scope of auditing has expanded from the examination of financial
statements to the inspection of how programs operate and how well they achieve their objectives, the concern remains largely the same: How does the “condition” observed in the matter under investigation compare to some “criterion” of performance?

When we look at the origins of program evaluation, on the other hand, we see a very different picture. Not only is evaluation quite recent in its major milestones, it counts many fewer practitioners than auditing and has had a much more mysterious development. In effect, some trace the origins of program evaluation to the Age of Reform in England, where studies of educational achievement conducted in the year 1870 brought contemporary recommendations for “incentives” to teachers in the form of “payment by results” (Madaus et al., 1983: 4). Others think history’s first evaluation took place earlier, when a steward of Nebuchadnezzar’s systematically tested the effects of a vegetable diet on the “fairness” and “fatness” of Daniel, Shadrak, Meshak, and Abednego (Old Testament, Dan. 1: 12-17, cited in Hoaglin et al., 1982: 7).

Whatever its origins, however, interest in the actual effects of government programs and policies—the principal subject of program evaluation—is hardly recent in the United States. The ability of evaluators to respond reasonably well to that interest, on the other hand, has developed remarkably over the past 20 years or so, as the gradual result of two apparently disparate and independent streams of intellectual inquiry.

One of these two streams was the effort begun during the 1950s to rationalize the management and resource allocation of defense missions and programs, an effort that eventually grew into the Department of Defense’s Planning, Programming, and Budgeting System (PPBS). This effort was only peripherally (if at all) concerned with verification, in the accounting sense; its main thrust was, first, to plan for program cost-effectiveness and then to evaluate (a) whether such cost-effectiveness had in fact been demonstrated, and (b) whether it was a result of the planning and programming process undergone. However, it is also true that only one component of PPBS involved the retrospective activities of program evaluation; rather, the major focus of the effort was on planning, so that the techniques it fostered—such as policy analysis (Wildavsky, 1966, 1969)—were essentially tailored by PPBS to delineate likely future effects rather than to identify the actual, observed effects of implemented, existing programs or policies. Developed largely by economists, these analytical techniques—policy analysis, cost-benefit analysis, cost-effectiveness analysis, systems analysis, operations research—all had economics as their core (Rhoads, 1978; Quade, 1975).
The second stream antedates the first in that it can be traced to agricultural research begun in the early 1900s. Experimental design and statistical analysis techniques were developed in an effort to find out which agricultural practices would lead to the best outcomes in terms of crop yield. These ideas were further developed by social scientists who contributed important advances in measurement. By the 1950s, large-scale retrospective evaluations were already being commonly performed, using methods of applied social research such as the survey and computer-assisted statistical analysis (Rossi et al., 1979). This path of evaluative development, however, was not focused on defense programs (as was the case for PPBS) but rather on social programs such as education, public health, or crime and delinquency. Nor was it the achievement of any one discipline; instead major methodological contributions were received from a broad array of fields including psychology, sociology, economics, political science, applied statistics, and anthropology.

Little by little, the two paths of evaluative inquiry have become less distinct, and today it is not uncommon to find a mixture of techniques from both streams used in a single study (Holahan, 1970; U.S. GAO, 1984). The yield has thus been an increasingly rich repertoire of methods for use in answering different types of questions about policies and programs.

Each of the two (auditing and evaluation) professions thus comes to program assessment via an extremely different disciplinary development and orientation. What this means is that although there are commonalities and overlaps between auditing and program evaluation—for example, both are retrospective, follow systematic processes, and are concerned with the usefulness of their work to its sponsors or to the public—there are also major differences in their definitions, purposes, the types of questions each is likely to address, and the manner of performing the audit or evaluation.

**DEFINITIONS, PURPOSES, AND QUESTIONS ADDRESSED BY AUDITING AND PROGRAM EVALUATION**

To begin, how are auditing and program evaluation defined today? The significance of an audit is given as follows by a 1980 text: "An audit means that an outside party is reviewing the information produced and the work done by management" (Edds, 1980: 8). A general definition of
Auditing, as developed by a professional accounting society, reads as follows:

Auditing is a systematic process of objectively obtaining and evaluating evidence regarding assertions about economic actions and events to ascertain the degree of correspondence between those assertions and established criteria and communicating the results to interested users [American Accounting Association, 1973: 7].

The same society also pinpoints the purpose of auditing:

The objective of auditing is to determine the degree of correspondence between that which is being audited, i.e., the auditee's operation and reports thereon, and established criteria [American Accounting Association, 1973: 8].

Thus, auditing is essentially normative, an examination of the match or discrepancy between a criterion (or standard or yardstick) and a condition (or the matter being audited). This approach, which clearly springs from the history of accounting and auditing, when applied to the performance of a government program, judges (or verifies) how well the program is doing against an existing norm.

A recent definition of program evaluation states that it is

the application of systematic research methods to the assessment of program design, implementation and effectiveness [Chelimsky, 1985].

The purpose of program evaluation is to provide empirical information that is useful to policymakers and program managers

- in developing and promulgating new programs (policy formulation);
- in assessing the operation of existing programs (policy execution); and
- in determining program effectiveness (accountability) [Chelimsky, 1985].

Although this purpose may involve the checking of correspondences between "what is" and "what should be," it is much more likely to involve methods allowing the empirical identification and measurement of what has, in fact, occurred. The focus of this effort is usually on the problem addressed by the program (its size, frequency, and direction), the experience of past programs (empirical findings and evaluation results), the characterization and systematic description of program
implementation (structured observation and interview), the estimation of what outcomes might have been expected in the absence of the program, and comparison of changes observed with those outcomes to determine actual program effects.

Three kinds of questions may be addressed by program evaluations: questions that are purely descriptive (these ask, for example, how many people are served by a program; or what are the characteristics of the populations served and of the program offered); questions that are normative (asking, for example, whether a program is operating according to congressional intent or executive branch regulation; or how changes associated with a program compare with the program's objectives); and questions that seek to establish cause-and-effect relationships (asking, for example, whether a particular environmental program has reduced pollution in a particular body of water; or whether a nutrition program has improved participants' health status).

In comparing the purposes of auditing and evaluation, then, it appears that auditing uses the relationships of what is found with what is generally agreed to be a correct or proper situation to report on the degree of correspondence between the two. That is, auditing asks a normative question, the purpose of which is essentially one of accountability. Program evaluation, however, in seeking to relate program activities to program effects in a way that will be useful for a broad array of information needs, asks other questions as well.

Both program evaluation and auditing, then, routinely address normative questions. Descriptive questions are also routine for program evaluation but are rare in auditing. Program evaluation alone, however, can address cause-and-effect questions. This is because the design used almost exclusively by auditors—that is, the criteria-referenced design—does not permit inferences about whether a program has caused the condition or changes observed. The reason this design does not permit causal inference is that it cannot produce an estimate of what the condition or change would have been in the absence of the program. Yet without such an estimate, competing causes—other than the program—for the condition observed cannot be ruled out. Program effectiveness evaluation does precisely that; that is, it rules out competing causes by estimating the extent to which observed changes, intended or unintended, are attributable to program activities. Cause is linked to effect through the use of design and analysis methods that compare conditions observed before and after implementation to estimates of conditions that would have been observed in the absence of the program. Different
evaluation designs are used to accomplish this, among which are experiments, quasi-experiments, and interrupted time series.

In summary, there are some noticeable distinctions between auditing and program evaluation that emerge from an examination of their definitions and purposes, along with the questions they address. First, auditing typically answers normative questions. Although program evaluation also does that, it additionally addresses descriptive and, especially, cause-and-effect questions. Another distinction lies in the assumptions posited by auditing and evaluation. The auditor assumes (via the audit definition) that there is a correct and agreed-upon way to do things and that "established criteria" can usually be found against which to measure conditions. The evaluator uses the criteria-referenced design more rarely, relies mainly on other designs, and typically focuses on the relationship between the changes that have been observed and the program, rather than the relationship between the changes that have been observed and the changes that should have been observed. This means that evaluation is much less affected than auditing by a lack of consensus on criteria or agreement on objectives. A third difference is that auditing does not produce estimates of what might have happened in the absence of the program being assessed, and evaluation does.

These are fundamental distinctions that influence, in their turn, the ways in which audits and evaluations are conceptualized, conducted, and judged for quality. Indeed, the fact is that when auditors and evaluators address the same type of question—or even the same question—their approaches and performance are likely to be very different.

CONTRASTING VIEWPOINTS AND EMPHASES OF AUDITORS AND EVALUATORS

Because of their dissimilar history, development, and orientation, auditors and evaluators tend to approach their work in contrasting ways. Important differences exist in their respective manners of

- treating and scrutinizing program objectives;
- examining program implementation; and
- determining program results or effects.
As might be imagined, these differences have effects across the entire execution of the audit or evaluation, which typically involves several different phases of work. These phases include

- formulating the project questions,
- designing the project,
- collecting data,
- recording and storing data, and
- analyzing data.

Each of these eight differences is examined below and shown in abbreviated form in Figure 1.

PROGRAM OBJECTIVES

Because of the auditor's concentration on a normative approach, the program objectives figuring in an audit need to be defined precisely. Because program success is typically measured in terms of the program's objectives, it becomes essential to the auditing process to pin down those objectives formally. Unfortunately, this is not always possible, so a major portion of the auditor's energy may be involved in trying to gain consensus on the objectives of a federal program against which to measure its success.

The evaluator also tries to understand program objectives, but the auditor's precision is not required as program effectiveness is not necessarily measured against program objectives. In addition, evaluation is interested in all actual outcomes of a program, whether intended or not by the program's formal objectives. Determining these effects cannot be accomplished using a normative approach; therefore the major question for effectiveness evaluation is not, "Did the program reach its objectives?" (to which the answer can only be yes or no, and if not, to what degree it missed—information that is more helpful for accountability purposes than for administrative learning and new policy formulation). Rather, the question is, "What would have been the results in the absence of the program, and—compared to that—what has actually been achieved?" regardless of program objectives.
<table>
<thead>
<tr>
<th>Assessment Dimensions</th>
<th>Contrasting Auditor and Evaluator Viewpoints and Emphases in the Performance of Program Assessments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Program objectives</strong></td>
<td>For A, it is essential to formally pin down stated program objectives because program success is most often measured against objectives. Hence, vagueness, uncertainty, or lack of consensus in the program's objectives is a major problem for the auditor.</td>
</tr>
<tr>
<td><strong>Program implementation or operations</strong></td>
<td>A's assumption is that if implementation is correctly performed (i.e., performed according to generally accepted yardsticks), this will be followed by the desired outcomes. A thus looks at how a program is supposed to operate (criteria), compares what is found (condition) to the prescribed criteria, and bases its conclusions about the quality of implementation on the correspondence between the two.</td>
</tr>
<tr>
<td><strong>Program results or effects</strong></td>
<td>A compares &quot;what happened&quot; in the program on selected outcomes with the program objectives for those outcomes and defines success in terms of the degree of correspondence between the two.</td>
</tr>
<tr>
<td><strong>Evaluation Emphasis (E)</strong></td>
<td>E tries to understand objectives but does not require auditing precision since program effectiveness is not necessarily measured against program objectives. E is more interested in measuring what happened as a result of the program (i.e., both intended and unintended effects) and in comparing that to what would have happened in the absence of the program.</td>
</tr>
<tr>
<td><strong>Program implementation or operations</strong></td>
<td>The correspondence between correct and actual implementation does not get a great deal of attention from E because E's assumption is that the relation between program implementation and observed changes in program data has to be demonstrated. E's interest will be most often in analytical description of program implementation and its variations, in the incidence of different operational characteristics, and in linking operational activities to changes in outcome data.</td>
</tr>
<tr>
<td><strong>Program results or effects</strong></td>
<td>E compares &quot;what happened&quot; in the program with what would have happened had the program not been implemented, and defines success in terms of the net effects of the program (i.e., those which are directly due to the program and would not have been observed absent its implementation).</td>
</tr>
</tbody>
</table>

**Figure 1**
| Formulating project questions | A's questions — based on the definition of auditing — are almost always normative. They are frequently not aimed at generalizations but rather at identifying deficits in a particular case or program. | E's questions are likely to be cause-and-effect (when the purpose is accountability, or when the need is to establish past or present effectiveness for policy formulation purposes), or descriptive as well as — more rarely — normative (for policy execution purposes). E's questions are often aimed at generalization (to improve future policy formulation and execution). |
| Designing projects | Since A's questions are most often normative, and its purpose most often accountability, its reasoning tends to be deductive and non-probabilistic. A's design of choice is the criteria-referenced design whether embodied in a case study or in a survey strategy. | E formally designs projects in terms of the kinds of evaluation questions posed (descriptive, normative, cause-and-effect) and in terms of the policy purpose to be served (formulation, execution, or accountability). E may reason deductively or inductively, but in the case of cause-and-effect questions serving an accountability purpose will typically design a project to support inductive (probabilistic) reasoning. |
| Collecting data or evidence | A is greatly concerned that evidence emanate from authoritative sources, and places heavy reliance on documentary evidence. A gives much attention to the verification of planned program procedures. Interviews and questionnaires provide supplementary information. | E places heavy reliance on structured interviews and questionnaires, and is concerned with the quality of the instruments used, and with the validity and reliability of the data collected. E's emphasis is on the compatibility of the data collection procedures with the design, on the one hand, and with the analysis to be performed, on the other. |
| Recording and storing data or evidence | A tends to record (and to archive) data in a linear, chronological way. This is related to the use of referencing as a quality-control mechanism. | E groups data to facilitate both the analysis to be performed, and the eventual re-use of the data (in meta-evaluation, for example, or in a capping report requiring re-analysis of the data). Grouping is according to theme or question answered. |

**Figure 1:** Contrasting Viewpoints and Emphases of Auditors and Evaluators
Analyzing data

Because of the normative, deductive character of auditing, a analysis tends to focus on individual cases, to search for critical events or discrepancies. However, in large-scale audits, many statistical techniques are currently being applied. These are not usually a formally planned part of the audit nor do they tend to make use of statistical probability.

Among the elements always included by E in a formal project design is the analysis plan. This plan is developed to provide the information needed to answer the question posed — taking into account the design, the sample, the type of data being collected and alternative analysis techniques. Data quality is carefully reviewed for validity and repeatability. E often generates estimates of events or conditions based on principles of statistical probability; uses statistical concepts like mean and variance; deals with data aggregated across many cases.
PROGRAM IMPLEMENTATION

Auditing tends to assume that if program implementation is correctly performed, it will be followed by the desired outcomes. The auditor therefore looks at how a program is supposed to operate, compares what actually happened to the prescribed procedures (i.e., standards or criteria), and concludes that any lack of correspondence between the two is a deficiency.

Evaluation assumes that the relationship between correct program implementation and any observed changes is a testable proposition. The correspondence between correct and actual implementation may not get a lot of evaluative attention. The evaluator is more interested in whether there is a relationship between actual implementation and observed program changes, the description of program implementation and its variations in different sites, or the incidence of distinct forms of program implementation. The interest springs from the desires of policymakers to "do something" and the need to understand why the evaluation results looked as they did.

PROGRAM EFFECTS OR RESULTS

Auditing draws upon legislative goals, federal regulations, or statements by executive branch program managers as expressions of program objectives. It then compares "what happened" on selected outcomes with the given objective and states results in terms of the degree to which the objectives were accomplished.

Although evaluation may adopt this normative auditing approach, more frequently it compares what happened with what might have happened in the absence of the program. This comparison allows an examination of the causes of what happened and permits attribution to the program of changes observed. Evaluation states results in terms of the degree to which the program had an effect, rather than the degree to which its objectives were achieved. If changes were observed but could not be attributed to the program, to an evaluator this would not be a program effect.

FORMULATING PROJECT QUESTIONS

As noted earlier, auditing questions are typically normative. They tend to focus on whether the program was properly implemented (as
compared against some yardstick—usually, generally accepted standards of implementation); whether program objectives were achieved; and, whether costs would have been lower given proper implementation. These questions are frequently aimed not at generalizations but at identifying individual deficits in achievement.

Again, evaluation tends to focus on descriptive questions (e.g., the amount and nature of variation in program implementation) and on cause-and-effect questions (e.g., the extent to which observed changes—intended or unintended—can be attributed to the program). Evaluation questions are frequently aimed at generalizations; that is, having determined that activity X produced effect Y, the evaluation will try to say how often and under what circumstances this relation is likely to be true.

**DESIGNING PROJECTS**

Auditing tends to design projects to support the deductive reasoning that follows from a normative approach. For example;

- All programs must have good management information systems (MIS) to be effective.
- Program X does not have a good MIS.
- Therefore, program X is not effective.

This reasoning is embodied in the criteria-referenced design, which is the design most often used by auditors, whether the strategy chosen is that of the case study or the survey.

Evaluation may also reason deductively but typically designs effectiveness projects to support inductive (probabilistic) reasoning. For example,

- Program effectiveness is positively correlated with the presence of a good MIS.
- Program X does not have a good MIS.
- Therefore, program X may not be (is probably not) effective.

Evaluators also use case studies and surveys as strategies but add to these some others, among which are field experiments and the use of available data. In general, evaluators designing projects are more likely to think in terms of the six approaches laid out in the Evaluation
Research Society's standards (i.e., front-end evaluation, evaluability assessment, process evaluation, impact evaluation, program monitoring, and meta-evaluation), which they key to the purpose of the evaluation—that is,

- policy formulation,
- policy execution, or
- accountability.

and to the type of question asked by the evaluation:

- descriptive,
- normative, or
- cause and effect.

COLLECTING DATA OR EVIDENCE

Auditors are greatly concerned that evidence emanate from authoritative sources; they place heavy reliance on documentary evidence. Auditing devotes much of its attention to verifying planned program procedures. Collecting information by questionnaire or structured interview is only a supplementary activity.

Evaluation places heavy reliance on structured interviews and questionnaires in anticipation of later statistical analysis of the data collected. Consequently, evaluators are greatly concerned with the quality of the instruments used and with the meaningfulness and repeatability (validity and reliability) of their measurements. Thus, formal instrumentation as well as interview structure and technique are matters of considerable importance to evaluators. Correspondingly, evaluators place much less emphasis than do auditors on planned procedures or their verification. The basic evaluation concern in data collection is that it be compatible with the design on the one hand and with the data analysis to be performed on the other.

RECORDING AND STORING DATA OR EVIDENCE

Auditors tend to record and store data in a linear, chronological way. This is important to the referencing process, which is the major control mechanism for ensuring the quality of auditing data or evidence.
Evaluators tend to group data with an eye toward the ensuing analysis and the demonstration of patterns and relationships in the data (e.g., all correspondents' answers to a given question are grouped together; pairwise groupings show a correlation between two variables). Evaluators are also concerned with the archiving of data to allow for reuse in meta-evaluation, for example, or in a capping report that requires reanalysis of the data.

ANALYZING DATA

As noted earlier, auditing reasoning is often based on deduction because of the normative character of the audit. The tendency is to analyze individual cases to search for critical events, discrepancies, and so on. The analysis does not typically make use of statistical probabilities, although statistical techniques of various kinds are currently being applied in large-scale audit surveys.

Evaluation reasoning tends to be inductive, often based on the statistical probability of an event or condition. Evaluators tend to deal with data aggregated across many cases and use statistical concepts such as mean and variance. A formal analysis plan is normally part of an evaluation design and ensures that at the end of the analysis, the information provided will answer the evaluative question(s) posed in terms of both relevance to the question(s) and other design criteria, such as generalizability and ruling out alternative explanations.

In summary, then, along eight dimensions (at least) of the work they do in assessing programs, evaluators and auditors are likely to take very different approaches, and there are probably many more differences than those outlined above. However, it is also the case that there are many similarities in the two types of approaches, these similarities give strength to the idea that the two professions can learn from each other. It was noted earlier that both approaches are retrospective, involve systematic processes, and focus strongly on relevance to users. In addition, both are involved today in assessing government programs and policies, and both are deeply concerned with the objectivity and independence of their work and its practitioners. Given these important bases for mutual understanding, two questions must be raised: What can the two professions learn from each other? And how should evaluators and auditors seek to shape the increasing methodological cross-fertilization now occurring between them?
POTENTIAL AREAS FOR INTERACTION BETWEEN AUDITORS AND EVALUATORS

Evaluators and auditors already rely on each other's work in many different areas. Evaluators have long borrowed accounting methods for measuring efficiency, especially in performing cost-benefit or cost-effectiveness analyses (Rossi et al., 1985). And this type of reliance can be expected to increase as evaluation places more and more emphasis on the front-end work needed for policy formulation and program planning. Auditors have also been borrowing survey instruments, increasingly using regression analysis and other techniques for examining data, and they appear to be coming to accept the idea of statistical sampling—albeit with occasional reservations because some auditors and accountants continue to feel that an audit needs to examine every transaction. However, this attitude is less often encountered in public sector auditing than it is in the private sector, and it is almost never encountered in that type of auditing known as "program results reviews" (that is, those audits that most resemble evaluations of program effectiveness). This, therefore, seems like a promising area on which to concentrate for future mutual advantage. Two dimensions of program effectiveness work seem to have immediate potential as areas in which program evaluators and auditors can exchange ideas and methods to their mutual benefit. These are the dimensions of cost and efficiency and project design.

COST AND EFFICIENCY

It was noted earlier that program evaluation's more recent origins could be traced to two sources—PPBS and applied social research—and that these two streams of intellectual inquiry had apparently moved closer together, at least in terms of the techniques currently employed in program evaluation. However, it is also the case that one stream proceeded gradually across the multidisciplinary fields and methods of applied social research, whereas the other experienced more explosive growth during the sixties, culminating in the federal development, implementation, and demise of PPBS (approximately 1960-1970). Although both streams pioneered or used and diffused methods (for example, surveys, case studies, experiments, operations research, longitudinal design, meta-evaluation) that have become part of the
everyday language of program evaluation, the sudden disappearance of PPBS circa 1970 meant a diminished emphasis within the field of program evaluation on both resource allocation and front-end planning. At the same time, the extent of the real difficulties involved in pinpointing the net impacts of policy interventions was becoming better known, so that few applied social research energies were available for focusing on costs and benefits (once those impacts had been determined), even if anyone had wanted to do so.

Today, however, program evaluation is becoming more rationalized as a function. Evaluation standards exist, and over the decade spanning the late sixties to the late seventies, federal funds spent on nondefense executive branch program evaluations rose from about $20 million to about $180 million. When the costs of defense evaluation and legislative branch evaluations are added, the figure moves to well over $1 billion for 1980. In addition, as Evans has phrased it,

The number of people who could lay reasonable claim to the title of program evaluators went from a handful to several thousands. The number of studies, the number of evaluation units in government agencies, the number of private research firms, the academic departments having programs in evaluation, the number of professional societies and journals devoted to evaluation, the number of legislatively mandated requirements for evaluation, and the actual use of evaluation findings by legislators [and] managers all increased dramatically during this brief period.

It therefore seems both reasonable and timely to suggest that evaluators start thinking once again about incorporating accounting perspectives into their work.

Evaluators should consider using cost-benefit and cost-effectiveness analyses—together with other evaluation techniques—for both policy formulation (front-end evaluation) work and accountability (effectiveness) assessments. The reason for doing this is that in choosing between new program alternatives, information on costs is critical. In some cases, when program evaluations have been performed for closely related programs, it is possible to develop front-end planning estimates of likely costs and benefits that are empirically based and hence relatively strong. But even when there is no empirical basis, it seems important, for sound policy formulation, to develop the planning estimates anyway, if only because “the discipline imposed by cost-benefit analysis forces the evaluator, policymaker, planner, and (program) manager to articulate economic considerations that might otherwise remain implicit or go unstated” (Rossi et al., 1979).
Again, with regard to program effectiveness evaluations, it is important to examine the costs and benefits involved to see whether the program was worth the investment. However, an important point needs to be made here. The analysis of costs should be done only after the net effects of the program have been established (according to whatever evaluation design may have been chosen), because cost obviously cannot be related to effects if effects have not been identified. This means that cost-effectiveness evaluation is an auxiliary, not an alternative, approach to effectiveness evaluation.

PROJECT DESIGN

Although evaluators have not, perhaps, given program costs the attention they deserve, it can also be argued that the difficulties of establishing the effects of programs may not have been taken sufficiently seriously by auditors. In comparing the procedures used by auditors and evaluators for determining program results or effects, there are notable differences in the way questions are defined and project designs are selected.

Program evaluators spend considerable time developing a formal technical design for their project. Such a design defines and poses the evaluation questions for study (whether descriptive, normative, or cause-and-effect); develops the methodological strategies for answering these questions; formulates a data collection plan that anticipates and addresses the problems and obstacles that are likely to be encountered; and details an analysis plan that will ensure that the questions posed are answered with the appropriate data in the best possible fashion.

The performance of program results reviews requires auditors to consider their designs carefully. Correlation analysis does not establish which of two variables causes the other, only that the two variables move (or don’t move) together. Before and after studies of the program taken alone cannot rule out causes other than the program. Therefore, it might be useful in the performance of program results reviews for auditors to begin using some of the techniques developed by program evaluators for designing assessments in which questions, strategies, methods, and analyses mesh tightly. The adoption of statistical techniques for estimating effects in the absence of the program would be especially helpful.

Although it seems clear that both professions could profit singularly from such an interchange, it is also true that philosophical as well as
methodological stances are involved. Auditing has been a normative process for several hundred years; proposing that auditors start now to answer cause-and-effect questions using inductive methods may be a difficult change to achieve. In the same way, evaluators have been more interested in program effects generally than in costs and savings. Given the relative youth of the evaluation profession and the evolutionary nature of evaluation practice, it may be easier for evaluators to incorporate cost considerations into their work than for auditors to alter their way of reasoning. Both changes are important, however. If either were to come about, it seems probable that major benefits would accrue in the performance of program assessment and in the consequent provision of useful information to program managers and policymakers.

REFERENCES


Eleanor Chelimsky is Director of the U.S. General Accounting Office’s Program Evaluation and Methodology Division. The division serves the Congress through evaluations of government programs, the development and demonstration of methods for evaluating programs, and the provision of design and measurement assistance to other GAO divisions. Mrs. Chelimsky came to the GAO from the MITRE Corporation. She is a former Fulbright scholar, past president of the Evaluation Research Society, and a member of the editorial review board for the Sage Research Progress Series in Evaluation. She serves on the editorial board of Policy Studies Review and was the recipient of the 1982 Myrdal Award for Government.