Do You Want a Performance Audit or a Program Evaluation?

Dwight F. Davis, State of Florida

Given the diversity of program evaluation, it is difficult to generalize about it, but certain major perspectives affect how evaluations are conceptualized, conducted, and reported. This article compares auditing with evaluation research in terms of foci, independence, quality control, and professionalism of program evaluation practitioners. It is important to note that the inferences drawn here represent tendencies rather than absolutes. Nevertheless, the theme is that legislators and executive agency managers should consider these tendencies before establishing program evaluation units and consider them again as they assess the performance of those units. If public policy makers do not carefully consider what they expect from program evaluation and incorporate those expectations into their requests for evaluations, then program evaluation practitioners will make these decisions for them according to their own theoretical and philosophical predispositions.

Definitions

Comparisons made here are not between two fields but, rather, between two traditions. Both auditing and evaluation research are concerned with assessing the worth of government actions. They differ in their theoretical and philosophical perspectives. As Olson observed with regard to the study of politics, the study of any particular phenomenon may vary according to researchers’ conceptual baggage.¹

Auditing here refers to performance audits. According to Generally Accepted Government Auditing Standards (GAGAS), performance audits involve a determination of the economy, efficiency, and effectiveness of government organizations, programs, activities, and functions, in addition to their compliance with laws and regulations.² Performance auditing evolved from accounting and financial auditing. Recent revision of GAGAS broadened the conceptualization of auditing, but the influence of accounting and financial auditing on the field is still strong and obvious. It is also important to note that auditing entities tend to be independent of those who manage programs. Indeed, this is a requirement of GAGAS.

Evaluation research means “the systematic application of social [science] research procedures in assessing the conceptualization and design, implementation, and utility of social intervention programs.”³ In many respects the theoretical foci of evaluation research are not significantly different from those of performance auditing, though the terminology is different. However, evaluation research evolved from social science, and its development has not been affected much if at all by those who are not social scientists. While much evaluation research work is done by people who are independent of managers of government programs, this is not always the case.

Why Is Choice Necessary?

At first glance, reasons for sorting out differences between auditing and evaluation research may not be apparent. As mentioned, both fields seem to be concerned with the same phenomena. It is instructive, however, that the literature in each tradition includes scant reference to literature of the other. This parochialism may be on the wane, but only in recent years. The point is that the his-

This article compares key features of performance auditing and evaluation research. Although these two traditions appear to be concerned with the same phenomena, important theoretical and philosophical differences distinguish them. Unless those who sponsor program evaluations are aware of differences among the various types of program evaluation practitioners and clearly specify what program evaluations should accomplish, a poor fit is found between program evaluation products and the information needs of public policy decision makers.

JANUARY/FEBRUARY 1990
Criteria for Establishing and Assessing Evaluation Units

Essentially two types of criteria are used to judge program evaluation work: technical quality and practicality. These criteria are interrelated but conceptually distinct. Technical quality refers to the nature of evidence gathered in a program evaluation. From a social science perspective, evaluation studies should have "construct validity, internal validity, external validity, and statistical conclusion validity." Auditors would say that audit evidence should be "relevant, material, and competent." In effect, both views provide that evaluation evidence should be certain to the maximum extent possible. Practicality of evaluation work refers to its timeliness, cost, interpretability, and responsiveness to decision makers' information needs, persuasiveness, and political sensitivity. One may add other dimensions, but these are the most commonly discussed.

Conflicts are integral to these criteria. In general, technical quality competes with practicality, especially with timeliness, cost, and interpretability. Some argue that technical quality requires independence of the evaluator from those who control what is being evaluated, but this may cause tension between technical quality and responsiveness to decision makers' information needs.

Tensions among criteria also characterize each type. For example, the control necessary for maximizing the certainty of causal inferences (internal validity) is in tension with the generalizability of results (external validity). On the practical side, it may be quite difficult to provide a timely study that is responsive to decision makers' information needs. What they want to know may take a long time to acquire. In addition, different decision makers (e.g., elected officials, agency managers, program managers) have different information needs.

The reason for mentioning the tensions among these criteria is to point out that comparing auditing to evaluation research is not easy. If all the tendencies discussed below are true and one uses the criteria mentioned above to compare these two traditions, decisions will still be difficult. Neither approach to program evaluation is clearly superior to the other on all criteria, and some characteristics of a field contribute to high ratings on a particular criterion while other characteristics lead to the opposite conclusion on the same criterion.

Comparing Auditing and Evaluation Research

Before comparing the characteristics of auditing and evaluation research, a few caveats are necessary. First, there have been no empirical comparative studies of these two approaches to program evaluation. Consequently, inferences drawn here are based largely on reading the prescriptive literature in both fields including published professional standards and textbooks. Inferences are also based on personal experience as a teacher of program evaluation and as a practitioner of performance auditing in the Florida Office of the Auditor General. Secondly, it is

It is unlikely that any particular program evaluation tradition can serve the information needs of all interested parties in public policy making. As Chelminsky has pointed out, there are several types of executive branch users and several types of legislative branch users of program evaluation. Each has somewhat different information needs. It follows that each should be enabled to gather the information needed, relying on the program evaluation tradition most appropriate to those information needs. Limited budgets do not allow for such analytical luxury. The trick is to fund the right type of program evaluation at the right time in the life of a public policy or program for the right audience.
likely that the tendencies mentioned here vary by context. Substantial variation exists in program evaluation contexts for both auditors and program evaluators. For example, auditing or evaluation research done on behalf of agency management is likely to vary from that done on behalf of a legislature. It also seems likely that auditing done by the U.S. General Accounting Office (GAO) differs in some important respects from that done by the much smaller, less well-funded, and often significantly younger state and local government auditing units. Although GAO is generally associated with auditing, the typical staff member is now called an "evaluator." Without doubt, the disciplinary backgrounds of program evaluation staff affect their approaches to program evaluation. Finally, in time, the distinctions made here may substantially blur as both auditing and evaluation research mature.

Comparisons made here concern four dimensions of program evaluation work:

1. The variables evaluators tend to emphasize.
2. Independence of program evaluators.
3. Quality control of program evaluation work.
4. Professionalism of program evaluation practitioners.

Each of these dimensions is multifaceted, and the discussion of them here is far from comprehensive. What is offered is a beginning for discussion of what sponsors of program evaluations want. Absent conscious decisions in this regard, the sponsors of program evaluations may not get what they expect.

Distinguishing Variables

Perhaps the most important distinction between auditing and evaluation research is that each tends to focus on different types of variables. Auditors tend to emphasize variables related to "management control."13 Such variables include organizational structure, plans, procedures, and monitoring; all of the activities central to the "principles of management" perspective on organizational life.14 Auditors are more likely to address program impacts now than they used to be, but they are still less likely to do so than evaluation researchers.

In contrast, evaluation researchers have tended to avoid management issues until the past 10 to 15 years. The emergence of the "program implementation" literature was in response to the tendency of evaluation researchers to focus on program impact measurement while treating programs as unexamined "black boxes."15 However, evaluation research interest in program implementation emphasizes the substance of government interventions and environmental influences rather than management control issues.16 The avoidance of program management issues by evaluation researchers appears much less likely now than in the past, but the difference in interest between auditors and evaluation researchers is probably still there nonetheless. Scott and Shore observed in 1979 that, in contrast with economists (and one might add auditors), sociologists tend to focus on variables that are not subject to control by governmental actors.17 This conclusion probably applies to other types of social-science program evaluators as well.

Auditors also tend to be more concerned with legal and procedural compliance issues than evaluation researchers. Auditors are obliged by GAGAS to examine such issues.18 No equivalent requirement is included in standards developed by the Evaluation Research Society. In other words, auditors appear to be more concerned with accountability than are evaluators. In part, this is probably due to the fact that, compared to evaluation research, audits are much more often done for interested third parties than for program management, but it is also a natural outgrowth of the concern with control in accounting.

The tendency to emphasize different variables has implications related to several practical criteria for assessing program evaluations. For example, given their emphasis on legal compliance, auditors are more likely than other kinds of evaluators to look for and detect fraud. Audits may be more responsive to the information needs of managers, because of audit emphasis on management control variables. A recent analysis of the use of U.S. General Accounting Office (GAO) recommendations supports this conclusion. Johnston attributed the success of GAO in getting its recommendations approved partly to GAO's tendency to emphasize changes that are managerial in nature rather than policy changes.19 This tendency toward narrower focus may also permit audits to be somewhat more timely than evaluations, particularly those that focus on broader, more complex issues. Assuming that legislators are mostly concerned with executive agency accountability, audits may be more valuable to legislators than evaluation research.

Not all arguments about scope of program evaluation favor auditing. The tendency of audits to focus on management control and compliance issues may result in program evaluations that are simply "nipping" or "bean counting."20 The emphasis on accountability may result in reports that are more critical in tone than necessary and, therefore, unbalanced or unfair. One authority on audit report writing urges auditors to write constructively, but notes that: "Balanced reporting versus exception reporting is still an open question for many professional experts."21 In addition, the emphasis on accountability may make auditors less likely to examine unintended consequences of government programs because such consequences are not necessarily related to legislative intent or formal statements of program goals and objectives. Such an oversight would be unfortunate because unintended consequences of
government programs may be at least as significant as those related to the official purposes of these programs. For example, the reduction of authorized speeds on U.S. highways in the 1970s was for the purpose of conserving fuel, but the continuation of this policy was typically justified in terms of reduced highway fatalities. In general, if legislators and other policy makers are mostly interested in policy advice, they are less likely to get it from auditors than from evaluation researchers, because such concerns go beyond management control and legal compliance.

**Evaluator Independence**

Auditors and evaluators are in general agreement that independence of evaluators from what they are evaluating enhances credibility of results. GAGAS requires that auditors avoid even the appearance of lack of independence.²² However, this view appears to be under revision within the evaluation research community.

Patton and others have argued persuasively that independence of evaluators has been oversold.²³ Strict evaluator independence increases the possibility that the evaluator will not address questions important to the user of the evaluation. It also assumes that evaluators do not have prejudices themselves, a human impossibility. At minimum, evaluators have methodological prejudices related to their disciplinary training. This is problematic because, as Patton explains: “(1) There is no single best plan for an evaluation, (2) there is no perfect design, and (3) there are always errors and ambiguities.”²⁴ If program managers and other interested parties are excluded from the design of program evaluation studies, there are no counters to the prejudices of the evaluators.

Even the standards developed by the Evaluation Research Society (currently, the American Evaluation Association) include the caution that: “Achieving a balance among potentially conflicting criteria will be a continuing challenge.”²⁵ It is further instructive that the first section of Evaluation Research Society (ERS) standards refers to “Formulation and Negotiation” during which “the concerned parties should strive for a clear mutual understanding of what is to be done, how it is to be done and why.”²⁶ (Emphasis added)

Strong independence of the evaluator from other parties interested in the evaluation may affect the timeliness of program evaluations as well as their scope and methodology. Since all evaluators, regardless of whether they are auditors or evaluation researchers, are concerned about the certainty of their results, there are always temptations to take as long as necessary in this regard. If evaluators are independent of influence by those for whom the evaluation is being done, evaluation work may not be timely in the interest of maximizing certainty of results. If this happens, the cost of the evaluation increases while its usefulness decreases.

Ultimately, the independence of an evaluator has more to do with legal protection against arbitrary firing than one’s philosophy, but how independence is exercised definitely is related to one’s philosophy. There appears to be a notable difference between auditors and evaluation researchers in this regard. It is difficult to assert this with confidence, however, since, in general, only auditors have been provided with protection from arbitrary firing. For example, it is not unusual for legislative auditors to be appointed to rather lengthy terms of office subject to removal only by extraordinary legislative action such as majority vote of both houses or impeachment. It is rare for the the head of an evaluation research unit to have such structural independence.

**Technical Quality**

No basis exists for arguing that either auditing or evaluation research is superior to the other with regard to technical quality. There is, however, an interesting difference between these two traditions in their approach to this issue. Auditing standards and textbooks include much richer and more detailed procedures for assuring quality work compared to the standards and textbooks of evaluation research. This is probably a natural result of auditor concern with control.

Whatever the reason, auditors appear to have much better guidance on how to implement quality control than evaluation researchers. For example, while both GAGAS and standards developed by ERS note the need to document evaluation work, only GAGAS prescribe key characteristics of “working papers” for this purpose.²⁷ Similarly, audit textbooks include detailed discussions of procedures for ensuring work quality throughout an audit.²⁸ In contrast, evaluation research textbooks tend to be light on procedural guidance though strong on the logic of research design.²⁹

The most recent revision of GAGAS requires that organizations conducting government audits have an “external quality control review” at least once every three years.³⁰ The purpose is to evaluate the organization’s quality control policies and procedures and their use. In essence, this is a check to see if the organization complies with GAGAS. No such requirement is included in the standards developed by the ERS. Although such a review does not necessarily involve an evaluation of the effectiveness of an audit organization in terms of its responsiveness to policy makers’ information needs or influence on public programs, it does constitute an important check on quality control.

The emphasis on quality control procedures in auditing standards and textbooks and advertisement of this fact by practitioners in the field may enhance the credibility of audits compared to evaluation research. Johnston’s findings regarding the success of the U.S. General Accounting Office in getting the vast majority of GAO recommendations accepted and implemented suggest this conclusion.³¹ Personal accounts of similar success of state legislative auditors further support this conclusion.³² It should be noted, however, that both the GAO and state legislative auditors benefit from the support they receive from their legislative bodies regardless of whether they have an...
auditing perspective or evaluation research perspective. For example, Johnston notes the extraordinary relationship between the GAO and Congress and the disadvantage of executive agencies in this regard.33

Although having well-developed quality control procedures is certainly beneficial, these procedures are not costless. If one views program evaluation as essentially a means by which government actors are held accountable to the polity through the legislature or other representatives, that may lead one to incur high costs for quality control of program evaluations. However, as decision-making research shows, it is rational to acquire additional information on a problem and how to solve it only if the cost of that information is less than the cost of uncertainty it reduces.34 In other words, there is a point at which the cost of reducing uncertainty of program evaluation results through quality control procedures is too high. Auditors may be more willing than evaluation researchers to cross this point. Spending too much on quality control will affect the timeliness as well as the cost of program evaluations.

Professionalism

Most evaluation researchers are not professionals who are devoted to a discipline in the same sense as auditors are, at least those auditors who are certified public accountants (CPAs) or certified internal auditors (CIAs). This does not mean that evaluation researchers are less ethical, reliable, or industrious than auditors. What differs is the means by which practitioners are developed and controlled. Most social scientists, including those who practice program evaluation, are not subject to disciplinary actions for noncompliance with professional norms as are CPAs and CIAs. Of course, one may be fired or have his or her contract cancelled for actions repugnant to those for whom the evaluation is being done, but this is a different kind of risk than professional sanctions. CPAs and CIAs are subject to having their licenses to practice revoked for noncompliance with professional norms. Furthermore, social scientists who work for such professionals must also abide by the norms of these professions because of the liability of their employers.

The difference in professional status of auditors compared to evaluation researchers has two important implications for the conduct of program evaluations by these practitioners. First, it appears that auditors may have higher credibility than evaluation researchers. Johnston’s analysis of the acceptance of GAO recommendations and the testimony of persons who work for state auditors suggest that auditors have more success than other types of program evaluators in getting their recommendations accepted.35 As Johnston hypothesizes, this may be related to the kinds of recommendations auditors make compared to evaluation researchers. However, it also seems likely that the certification of many auditors also enhances their credibility. The assumption that certification leads to appropriate conduct is the foundation of all professional regulation by government. If auditors are more credible than evaluation researchers, the work of auditors may be more useful during bargaining and negotiation that ensue during public policy deliberations.

A potential drawback of professionalism is the tendency to place high value on uniformity and consistency of behavior among practitioners within a given profession. This is this feature of professionalism that reassures those who buy professional services that they are getting competent service. This characteristic, however, may also lead to conceptual ossification.36 Innovation and change involve risks for professionals. Any behavior which could be interpreted as inconsistent with professional consensus makes a professional potentially liable for negative results. Consequently, professionals are cautious about adopting innovations.

There are reasons for believing that auditing is less innovative and conceptually rich than evaluation research. One measure of this difference is the relative volume of literature in these two traditions. An awesome volume and diversity of evaluation research literature exists, but the performance auditing literature is scant by comparison.37 Several journals are devoted exclusively to evaluation research, but no journals are devoted to performance auditing.38 With the exception of accounting and file review methods which evolved from financial auditing experience, the methodological innovations in program evaluation have come from evaluation researchers. Some of the controversy surrounding the revision and adoption of GAGAS has largely concerned evaluation researchers’ fears that these standards are too closely related to financial auditing practice and are, therefore, inapplicable to many program evaluation situations.39 On the other hand, many auditors were concerned that the most recent revisions of GAGAS may go beyond the boundaries of auditing.40 If, as GAO found, many auditors have had difficulty complying with GAGAS in the past, it seems likely that the more expansive view of auditing expressed in the 1988 edition of GAGAS will cause further problems for the profession.

If the professionalism of auditing results in auditors being more cautious than evaluation researchers, an interesting choice dilemma emerges for those sponsoring program evaluations. One may employ an auditor with reasonable assurance that he or she will use well tested, conventional approaches to the work but at the risk of avoiding questions which cannot be addressed with these approaches. Alternatively, one may employ an evaluation researcher who may be more willing to take risks in the interest of addressing questions of interest to the program
evaluation sponsor. In general, the more material the issue, the less likely it is that it can be addressed with certainty using conventional approaches.

What Type of Program Evaluation Do You Want?

When deciding to establish a program evaluation unit, requesting work from such units, or contracting for evaluations, policy makers cannot assume that no important differences distinguish the various kinds of practitioners in this field. Program evaluation practitioners, like other professionals, are affected by their disciplinary backgrounds and the traditions within which they have worked. This article highlights some apparent differences between auditing and evaluation research. One might easily argue that greater variations are evident within program evaluation than just these two traditions. The point is that, because varied approaches to program evaluation exist in fact, practitioners tend to give policy makers whatever their program evaluation traditions allow, not necessarily what policy makers want or need.

Policy makers must decide what they expect from program evaluations. Policy makers need both process/managerial data and impact/outcome data for full evaluation of policy alternatives and priorities. Program evaluations should also address different concerns at different stages of policy development. To compare the options for providing program evaluation services, policy makers need to know what information is currently lacking and then proceed to acquire that information. This is not an academic problem. Cheilmsky, Cordray, and Datta recently observed that federal policy makers are not being provided with the full range of relevant policy evaluation information. It is not clear whether this is the result of conscious choice by federal policy makers or the result of some interests being more influential than others during a time of budgetary cutback.

Probably the best that program evaluation sponsors can do is to be as explicit as possible about what they want to find out through particular program evaluations and also about what kinds of evidence they will accept. This implies that program evaluation scope and methods must be negotiated. Absent such negotiation, the fit between policy makers' information needs and program evaluation results will continue to be problematic.

Dwight Davis is an Analyst for the Florida Department of Highway Safety and Motor Vehicles and an Adjunct Professor of Public Administration at Florida State University. He has served on the faculties of the University of Oklahoma and Texas A&M University, and he spent seven years supervising performance audits of state government programs in Florida for the Florida Office of the Auditor General.

Notes


13. For examples see Leo Herbert, Auditing the Performance of Man...


16. For example see the discussion of “why programs fail” in Rossi and Freeman, op. cit., ch. 4; and Mary Ann Scheier, “Program Theory and Implementation Theory: Implications for Evaluators,” New Directions for Program Evaluation, no. 33 (Spring 1987), pp. 59-76.


25. ERS Standards Committee, op. cit., p. 10.

26. Ibid., p. 12.

27. U.S. General Accounting Office, op. cit., pp. 6-21 and 6-22

28. For example see Malan et al., op. cit., ch. 5.

29. It is instructive that one of the best selling program evaluation textbooks does not include a chapter on managing program evaluations. See Rossi and Freeman, op. cit.


32. Such accounts have frequently been heard at meetings of the Legislative Program Evaluation Section (currently the Legislative Program Evaluation Society) of the National Conference of State Legislators.


35. Johnston, op. cit.


37. One indication of the tremendous volume of evaluation research literature is that one publisher (Sage Publications, Inc.) published a two volume Handbook of Evaluation in 1975 amounting to approximately 1,600 pages, and every year since has published a lengthy Evaluation Studies Review Annual, which includes important writings in the field, many of which are reprinted from other sources. In addition, the many textbooks in evaluation research typically contain substantial bibliographies. In contrast, as recently as 1984, the authors of a major auditing textbook observed that: “Although performance auditing has been implemented effectively for well over a decade, surprisingly little literature has been published on the subject.” Malan et al., op. cit., p. 1.


40. Personal communication from Bernard Geizer who served on the panel established by the U.S. General Accounting Office to consider proposed revisions to Generally Accepted Governmental Auditing Standards during the 1988 recent revision.

41. Chelimsky, Cordray, and Datta, op. cit.