DOCUMENT RESUME

ED 386 462 TM 023 798

AUTHOR Wilkinson, L. David; Waring, Colleen G.
TITLE Evaluation and Performance Auditing: A Rose by Any Other Name.
PUB DATE Apr 95
NOTE 35p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Causal Models; *Cooperation; Cost Effectiveness; Elementary Secondary Education; *Evaluation Methods; *Evaluators; Policy Formation; *Program Evaluation; Public Policy; School Districts; Urban Areas; *Urban Schools; Urban Youth
IDENTIFIERS Auditors; Austin Independent School District TX; Drug Abuse Resistance Education Program; *Performance Based Evaluation

ABSTRACT

This paper presents a comparison of performance auditing and evaluation. It is the product of a collaboration between the senior evaluator from the Office of Research and Evaluation in the Austin Independent School District (Texas) and the senior auditor of Austin's City Auditor's Office. These officials were required to work together when the city undertook an audit of the Drug Abuse Resistance Education (DARE) program. The working relationship continued under the auspices of a partnered audit of social policy issues associated with the city's youth. It is concluded that evaluation and performance auditing are, in their best expressions, very much alike. One notable difference is that auditing cannot address cause-and-effect questions, which concerns program effectiveness evaluation. Both professions could benefit from each other's methodology and viewpoints. Particularly, in the study of educational impact, closer cooperation between school district evaluators and auditors may serve to leverage their resources and maximize their information processes. Two tables summarize the comparisons. (Contains 33 references.) (SLD)

EVALUATION AND PERFORMANCE AUDITING

A Rose by Any Other Name

L. David Wilkinson

Department of Performance Auditing and Evaluation
(formerly Office of Research and Evaluation)
Austin Independent School District

Colleen G. Waring, CIA

Office of the City Auditor
City of Austin

Austin, Texas

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1995

DISCLAIMER

The opinions and conclusions expressed herein are those of the authors and do not necessarily reflect the position or policy of the Austin Independent School District or the City of Austin. No official endorsement should be inferred.

"What's in a name? that which we call a rose
By any other name would smell as sweet."
(*Romeo and Juliet*, Act II, sc. ii, 43)

INTRODUCTION

This paper presents a comparison of performance auditing and evaluation. It is the product of a collaboration, which grew out of a professional collaboration in which two people from different backgrounds and disciplines were compelled to build the bridges necessary to work together. This paper describes the understanding which we think we achieved.

PERSPECTIVE

The authors of this paper are, respectively, the senior evaluator in the Office of Research and Evaluation (ORE) in the Austin Independent School District (the Austin, Texas public schools) and the senior auditor in the Office of the City Auditor in the City of Austin, in which the school district operates. The authors and their organizations had occasion to work together beginning in the 1993-94 school year when the City Auditor's Office undertook an audit of the Drug Abuse Resistance Education (DARE) program, which is primarily funded by the City, through its police department, supplemented by the funding from the District, through federal Drug-Free Schools and Communities (DFSC) grant monies. This working relationship continued under the auspices of a "partnered" audit of social policy issues associated with the city's youth.

As professionals working together, we had first to get past the initial strangeness of different vocabulary, different work routines, and even a tinge of some interorganizational history. After we had dealt with these superficial considerations, we discovered a more profound need to understand one another's thinking as it derived from our different backgrounds, training, and experiences. Much seemed mutually familiar, but some aspects of the other's field seemed less apprehensible.

Now, having spent a year and a half working together to attain a better understanding of one another's fields, we believe other school districts and other governmental entities might profit from our experience, particularly because the governing bodies of governmental organizations are requiring the kind of efficiency and effectiveness information both evaluations and performance audits provide. We also believe that practitioners of our respective fields would profit from a self-examination of their assumptions and beliefs about the types of investigations they conduct.

OBJECTIVES

1. To describe the experience of working collaboratively on a "partnered" audit, touching on some of the difficulties and rewards of professionals from different backgrounds and disciplines working together;

2. To compare and contrast the philosophy, methodology, and practice of evaluation and performance auditing;

3. To describe the commonalties and differences between evaluation and performance auditing as they are applied in the governmental sector, with an emphasis on the relationship of the fields to their published standards; and

4. To define terms, debunk misconceptions, and create a better understanding of performance auditing.

ORGANIZATIONAL AND PERSONAL BACKGROUNDS

Our collaboration was facilitated because of the characteristics of the organizations and the personal background of the authors.

The Austin City Auditor's Office–On the Cutting Edge

Neither Austin's Office of the City Auditor (OCA), nor the senior auditor overseeing the DARE audit, are fully representative of their profession. Having begun in 1984 to introduce *performance audits* into the Austin internal auditing environment, the office positioned itself with a very few local government audit shops at the front of an evolving trend. It must be recognized that performance auditing is not limited merely to program economy and efficiency objectives. The scope of the OCA's performance audits incorporate and combine all five of the major objectives of auditing:

- Safeguarding of assets;
- Compliance with relevant laws and regulations;
- Reliability of performance and financial information;
- Efficiency and economy of operations; and
- Accomplishment of goals and objectives (*effectiveness* as defined by the audit model).

The OCA's rationale for encompassing all five categories in its performance audits is based on the fundamental principle that *performance* of an organization includes *all* aspects: whether they are associated with accomplishing the mission or with maintaining financial viability.

OCA differs from most local government internal audit groups, principally in the extent to which the office's resources are allocated to conducting performance audits. The staff believe that their work places OCA at the front of professional evolution in the internal auditing field, because performance audits have the potential to add greater value to their environments than a narrow focus on internal controls. Where the typical local government internal audit office concentrates its efforts on evaluating the adequacy of internal financial controls, OCA's concentration on performance audits further distinguishes it from the mainstream by requiring a wider variety of staff specialties and a larger number of staff. Moreover, there are considerable differences in the size and style of audit reports, among other things. Although OCA is not alone in its attention to performance of government operations, the investment required to produce a performance audit has prevented many small internal audit shops from jumping on the bandwagon.

Evolution of Auditing

The increasing interest in performance auditing at the local government level was documented in a 1986 survey of 750 local government (city, county, township, school districts, and other small jurisdictions) budget practitioners about the use of performance auditing in their jurisdictions. Of 524 respondents to the survey, 32 percent were using performance auditing (Stipak & O'Toole, 1990). Another study, carried out in 1987, surveyed the municipal finance officers of the 152 cities with populations over 100,000. Of the 170 replies, 12.9 percent reported regular use of economy and efficiency audits, and 7.9 percent reported regular use of program effectiveness audits (Parle, Wallace, & Davis, 1990).

Today, 11 years after the initial introduction of performance auditing in OCA, the profession of government internal auditing continues to ponder the necessity for changing from the traditional focus on financial transaction controls. Nearly every issue of the profession's bimonthly membership journal, The Internal Auditor, contains at least one article making a clarion call to auditors to heed the winds of change, and recognize the demands of their customers (both management and policy makers alike) for the internal auditing office to begin to "add value" (Wernz, 1994; Ratliff, 1994; Julien, 1993; Paape, 1993; Flaherty & Stein, 1991; Burns, 1991; Thompson, 1991).

Ultimately, *adding value* in an auditing environment translates to converting from the traditional focus on transaction controls to the full range of performance auditing. The City Auditor of Oakland said it best: "If we commit our limited internal audit resources to counting petty cash or endless efforts to see whether every insignificant document has been properly signed, we are betraying the public trust" (Ng Lau, 1994). However, such a change does appear to be taking place incrementally–more and more of the audit reports we receive from our colleagues have begun to reflect a focus on other aspects of operations besides transaction accounting and control. In an article on "reinventing" the auditing profession, McNamee and McNamee (1992) said:

> "Auditing began by observing and counting, or reperforming, the work of others. This practice lasted for nearly 5,000 years, or until 1941, when Victor Brink introduced the concept of systems auditing, which focused on audits of system controls, rather than on checking transactions. This was a dramatic change in internal auditing that still has not been adopted in some countries of the world.
>
> "Changes in the environment are pulling us toward another breakpoint. Audit in the 21st century–on the other side of the breakpoint–will be very different from what it is today.
>
> "Audit is the process of comparing *what is* to *what should be*. This clearly defined purpose will continue to be true; it will be our anchor as we innovate, reinvent, and begin a new growth cycle....
>
> "...Given what we know about future pull, about creativity, and connections, and what we know about the principles that underlie successful growth, we need to find a different source of "what should be." That source is *the shared vision of the organization*."

Despite this dawning recognition of the evolutionary pressures that are forcing change in the scope and objectives of audits, many internal audit offices have not yet begun modifying their recruiting practices to broaden the backgrounds of their staffs (Malan, 1991). This stability in the recruiting practices is evidenced by data from the biannual job market survey of internal auditing departments conducted at the University of Arkansas Department of Accounting. The 1994 survey found that 68 percent of audit directors reported that they are recruiting staff with an accounting or auditing background. The authors note that "These percentages have changed very little over the past several years" (Oxner & Kusel, 1994). In fact, the 1992 salary survey by the same group noted that 63.8 percent of respondents "most desired" an auditing or accounting background for internal auditor candidates (Kusel, 1992). The 1992 report also noted that 70.7 percent of male auditors and 72.6 percent of female auditors held undergraduate accounting degrees.

Other information also seems to support the perspective that the internal auditing profession is disproportionately peopled by accounting types. Of 21 U.S. respondents to a survey comparing

U.S. audit functions and characteristics to those of Japanese audit groups, two U.S. audit departments reported requiring 3-4 years in accounting for employment, while 12 reported "business experience in area" as their requirement. While at first glance these results appear to indicate a broadening of experience requirements from the old narrow focus on accounting, additional responses to the survey reveal that four of the respondents require their hires to possess the Certified Public Accountant (CPA) designation, and an additional 10 indicated that they require both the CPA and a Certified Internal Auditor (CIA) designation (Burnaby, Powell, & Strickland, 1992).

The linkage between an accounting background and the types of audits carried out is not inconsiderable. The initial six hours of beginning and introductory accounting focuses almost entirely on transactions. Although subsequent mid-level course work begins to develop the students' understanding of more theoretical principles, the final stage of this course of study draws the student back to applying the theoretical principles to specific individual transactions. The early coursework's focus on minute detail has a screening effect, weeding out most students whose personalities are unsuited to the exacting, unremitting focus on individual transactions and control procedures.

Neither the demographics of the OCA, nor the background of the senior auditor assigned to supervise the DARE audit, share the industry's plurality of accounting experience. Of 22 full-time auditing professionals, eight are CPAs. Of these CPAs, five are somewhat unusual in their backgrounds, in that their CPAs represent a mid-life career change. One was an English teacher; another originally obtained her Ph.D. in physics, and taught high school science for 17 years; another began his working life with a B.S. in chemistry, and worked as a quality control manager in a large paper company. One obtained his B.S. in communications while a technician in the Air Force, and the last of the five also has a law degree.

The Auditor In Charge of the DARE audit began her career in the communications industry, working for a daily newspaper, a publishing company, and several advertising agencies. She gained her interest in auditing while working as an editor of audit reports for the now-defunct Legislative Fiscal Office of the Oklahoma Legislature. After five years of editing other auditors' reports first in Oklahoma, then in Austin, she was finally able to convince management of OCA that she could conduct an audit. Six years and a dozen performance audits later, she found herself assigned to lead the OCA's second year of Opportunities For Youth audits. Among this formidable group of social service program performance audits lurked the politically charged DARE audit, which led to the Austin Independent School District (AISD) and its Office of Research and Evaluation (ORE).

The Office of Research and Evaluation

AISD made a commitment to research and evaluation 21 years ago with the formation of ORE. The mission of the office is to provide objective, accurate, and timely information to decision makers. The information can range from an individual student's test scores to evaluation reports on instructional programs, and decision makers can be as different as a parent concerned about a child's achievement and a federal funding agency.

Originally begun with experimental federal grant funds in 1973-74, ORE became the District's internal evaluation organ, employing both local and grant-funded staff. ORE conducted comprehensive evaluations of federal assistance programs--Title I, Title IV, and Title VII--as well

as District initiatives such as early ventures into the quarter system and individually guided education. In 1976, ORE took over the District's fledgling testing program, expanding it into a systemwide vehicle for program evaluation and school accountability.

Through the years, ORE built a national reputation for the quality of its work, particularly in the areas of dropout prevention, retention, and methodological innovation. Two of its former directors were officers in Division H of AERA, and more than 20 of its reports have won in the annual outstanding publications competition held by Division H.

Over an 18-year career with ORE, the senior evaluator has conducted and supervised dozens of evaluations of many different kinds of programs in the areas of compensatory education, special education, bilingual education, vocational education, gifted education, staff development, drug-free schools, dropout prevention, and dropout recovery. Five of ORE's publications awards bear his name.

Ironically, during the course of preparing this paper, the District hired a new superintendent and underwent an administrative reorganization. For both budgetary and political reasons, ORE was dissolved and its functions split among several departments. The evaluation component was merged with the Department of Internal Audit to become the Department of Performance Audit and Evaluation. Because the collaboration which stimulated this paper took place when ORE was still intact, it is simpler and less awkward to refer to ORE in the present tense, even though, in fact, it no longer exists. It is hoped that the new department will effect a productive synthesis of its two "houses."

METHOD

COLLABORATIVE WORK

We began the process of mutual and self-understanding undramatically, simply by talking. The City Auditor's office had been charged with an audit of the Drug Abuse Resistance and Education (DARE) program, and the auditors had in mind utilizing such academic performance measures as the District had in place, as well as adding a dimension beyond that which the District had employed: an examination of the impact of the program on juvenile crime. At an introductory meeting, we quickly discovered that, though we shared a mutual interest in investigating the effects of DARE, we were speaking a different language, approached the task from different directions, and had some interorganizational baggage to dispose of as well.

Different Vocabulary

In our first few months together, our interaction might be described as a sort of "he said, she said" dialogue. The following exchange is fictitious (and more grammatical than real speech), but it somewhat captures the flavor of the interaction:

SHE: "In the survey phase..."

HE: "Huh? Excuse me. Whom would we be surveying? What items would be on the questionnaire?"

SHE: "Questionnaire? I'm sorry, we use the word **survey** to refer to the initial phase of the audit when we scan the 'landscape' of our auditee environment to get an idea of the most significant risks and try to assess the auditee's vulnerability to each risk."

HE: "Risks? Do you actually think this assignment will be dangerous? I know DARE is taught by uniformed police officers, but I assumed that they wouldn't try to shoot us..."

SHE: "Ooops, '**risk**' is an auditing term for the kinds of things that could go wrong, sort of the inherent consequences associated with the specific operation we're auditing. Like in our audit of the parks department, we learned that customer and employee safety was one of the significant risks of that operation."

HE: "How does the concept of risk relate to an audit of DARE?"

SHE: "Well, we have been asked to evaluate the effectiveness of the DARE program, so I would define the risk we are looking at as 'the possibility that the program might not fulfill its mission.' What we do in the survey phase of the audit is try to assess, in a superficial way, how likely it is that that risk might actually occur–how vulnerable the DARE program is to the risk of not meeting its mission."

HE: "It almost sounds like a way of stating the null hypothesis. But why is it necessary to determine the probability of the program's not achieving its mission? Isn't that what you've already decided to evaluate?"

SHE: "Because auditing is such an expensive activity, if we determine early in our work that the vulnerability is very low–in other words, that the probability of the DARE program not meeting its mission is low–then we would want to revisit the need for this audit, or determine if some other vulnerability is higher, and consequently more important to audit.

"In addition, we also use the term '**survey phase**' to stand for other preliminary tasks besides the risk and vulnerability assessment. We have to learn about the program, get the background and history, find out what kinds of performance measures are currently being tracked, what other studies or audits have already been done, and whether we can rely on that other work. It could be called the 'environmental scan.'"

HE: "Oh, I see. I think what you're talking about is the evaluation plan. So you use the survey to develop the evaluation plan?"

SHE: "We use the survey portion of the audit to decide what our audit plan will be. At the end of the survey, we will establish the O, S, and M."

HE: "S and M? I thought we were talking about auditing."

SHE: "That's **objectives**, **scope**, and **methodology**. We set out what we are trying to accomplish and how we are going to go about doing it. Now, in the field work phase of the audit..."

HE: "Excuse me again. What we've been talking about so far is obtaining data from extant computer files. What data are going to be collected in the field? Are you talking about interviewing people, or are we still talking about surveys?"

SHE: "If you imagine the survey phase to be a broad, but shallow, look at the landscape of our audit subject, then the **field work** phase is when we get down to the deep digging in a specific area. Our survey of the whole tract tells us where is the best place and what is the best method to get down to bedrock, or the 'bottom line,' as the City Council likes to call it."

HE: "Oh, I get it. That's what we could call the data collection phase of an evaluation."

"The term 'audit' confuses me. I think of an audit as having to do with money, but what you're describing sounds to me like an evaluation study. You are going to be collecting data and doing analyses. Could we just call it a study?"

SHE: [appalled] "Not if we are going to conduct this audit in accordance with Generally Accepted Government Auditing Standards. (We usually call them 'GAGAS,' because sometimes the amount of work required to comply with GAGAS does. –That's an auditor joke.)"

HE: Then what's the difference between an audit and a study?

The reader gets the picture. When we were not talking at cross purposes and stumbling over new terminology, we were trying to get a shared view of what work was going to be done and how. Each of us was so accustomed to, and comfortable with, the vocabulary of her or his respective field that the other's "assignment of conceptual space" (Sternberg, lecture) seemed foreign, even perverse.

As it turned out, that last, seemingly innocent question launched the continuing discussion that ultimately gave birth to this paper. We struggled to find a definition of audit that differentiates it from evaluation in order to explain the practical differences between our work. During our discussions we encountered differences in how the work is carried out, the customers' perceptions of the work product, and even a fundamental difference between the ultimate missions of the two practices.

Different Work Routines

Another difference which complicated our working together was in our work routines, particularly in regard to data collection and data analysis. At the time, ORE had the luxury of taking an "exploratory" approach to data gathering and analysis, a characteristic research and evaluation approach which auditing does not share. When the City Auditor's Office apprised ORE of its time line for the audit, the reaction was one of amused disbelief. It was not so much that ORE does not work under stringent time lines--indeed, new evaluators used to a graduate school routine find themselves working at what sometimes seems a breakneck pace--but that the time line did not seem to allow sufficient time for reflection about the data and for additional inquiry that might be suggested by a preliminary analysis of the data.

ORE typically proceeds by identifying the data needed to answer questions of interest, then specifying in detail data processing and data manipulation tasks to be accomplished by mainframe programmers. In AISD's mainframe computer environment in which student and other data are maintained on multiple, nonrelational data files, data extraction and file manipulation often require extensive programming through the use of Statistical Analysis System (SAS) and sometimes

COBOL. ORE relies on its access to mainframe data files and programmers and is accustomed to specifying data processing tasks for mainframe programmers, analyzing the output, setting out additional analyses, and so on in an iterative process. OCA has very limited access to computer programming resources and so does not set out tasks which require iterative analysis.

These differences, which we did not appreciate fully at the time, stemmed from the pivotal issue, previously noted, of audit versus evaluation study. What was an audit to the senior auditor, meaning that the audit should move along in certain, predetermined steps, was a study to the senior evaluator, implying a more open-ended, and hence more time-consuming, process. In short, while the evaluator was thinking of the audit as just another in an indeterminate series of studies about the issue of student drug use, focusing on one possible intervention (the DARE program), the auditor was thinking of the audit as a one-time endeavor.

Interorganizational History

Another issue complicating our collaboration was our interorganizational history. The City and District, as independent, though interrelated governmental entities, coexist in a somewhat uneasy truce, against the backdrop of a highly educated, activist citizenry, and a frequently adversarial press. Any initiative in an arena regarded by the other as belonging in its province is viewed with some apprehension, the more so because it may find itself portrayed unfavorably in the media.

Stereotypic attitudes in the District hold that the City is always throwing its weight around and taking action without regard to the District. For example, City building codes have some notoriety in the District. Even a portable classroom, a temporary structure, has to meet City codes. The City's attitude is that all structures, even District buildings, come under its provenance. This has created a considerable amount of conflict with respect to the City's considerable environmental codes. The District, as an independent governmental entity, believes that it can be trusted to meet the environmental remediation requirements. However, the City, having experienced the District's ponderous timing in coming into compliance with these laws, has insisted that the District put up the same escrow funds which are required of all private developers. The amount of the funds involved in these disputes is large enough to create considerable conflict among the parties.

For its part, the District is regarded by governmental and other entities in the city and county as so insular as to be virtually unapproachable for any sort of collaborative endeavor. Attempts to involve the District in collaborative projects are met with a half-hearted, disjointed response, often accompanied by the plaint, "You don't know what we're dealing with." In response, the City often pursues its own agenda in headlong fashion (as the District sees it), compelling the District to cooperate or be embarrassed in the forum of public opinion. A recent instance involved the City directing the police department to pick up juveniles who were not in school and deliver them bodily into the custody of the District. The District barely had time to react to this proposal before it was implemented and before it received extensive favorable media attention, despite District concerns about the detention of students legitimately on their way to part-time jobs or on other authorized activities (e.g., student journalists), students from other local school districts abroad in the city, expelled students, and dropouts, apart from such procedural matters as where students should be delivered, how they should be processed before return to their campuses, whether to try to hold students unwilling to return to school, whether to call students' parents, how to transport students back to their schools, and what to do with students until they could be returned to their schools. These issues notwithstanding, the City went ahead with its plans, leaving the District scrambling to respond as best it could.

Given these sorts of interoganizational dynamics, it may be rightly suspected that the District was wary of entering into another City-sponsored initiative, especially one as potentially inflammatory as an audit of the popular DARE program. Nonetheless, we had to learn to trust one another and to take the other on faith. As previously related, our organizational and personal backgrounds made this possible. Indeed, as it developed, the DARE audit proceeded largely as an informal effort between our offices that was never documented in writing. The senior evaluator and senior auditor, with tacit approval from their supervisors, took a great deal on themselves, even to the extent of accomplishing a data match through a third-party agency which could have gone aground at several points because of bureaucratic strictures and confidentiality concerns. Both offices were actively interested in determining the effectiveness of the program--ORE from curiosity and a desire to undertake an analysis not included in its six years of program evaluations, and OCA to carry out its charge as comprehensively as possible.

REVIEW OF THE LITERATURE

When we began to go beyond the superficial differences in our work habits and work environments, we set out to understand the differences in our respective fields from a more formal, academic perspective. To this end, we conducted a review of the literature to ascertain whether comparisons of performance auditing and evaluation already existed. A few, but very informative, studies address the similarities and differences between performance auditing and evaluation. Noting the "remarkable paucity of literature on this subject," Chelimsky (1985) describes the pioneering work of the General Accounting Office (GAO) to bring auditing and program evaluation under the same roof. She examines four aspects of the relationship between auditing and evaluation: (1) their different origins; (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. She notes some fundamental differences between auditing and program evaluation, but she concludes that major benefits could accrue if an interchange in methodology could be accomplished.

Davis (1990), writing from the perspective of a policy analyst, compares the key features of performance auditing and evaluation research--foci, independence, quality control, and professionalism of practitioners--and distinguishes important theoretical and philosophical differences between them. He develops the theme that legislators and agency managers should recognize the differences between the two traditions in deciding whether program evaluation products fit the information needs of public policy decision makers.

Besides the literature review, we also consulted our professional and personal libraries. Relevant texts and other reference materials are cited in the bibliography.

ANALYSIS OF STANDARDS

In comparing the differences between performance auditing and evaluation, we determined that a major source for reference was the standards which governed our respective fields. We found that, in trying to explain to each other why we would do something in a certain way, we would often cite our respective standards. Therefore, we undertook a critical examination of those standards to ascertain commonalties and differences.

There are a number of standards which apply to the conduct of audit and evaluation work. We selected for analysis the most applicable standards of our professions:

- The Program Evaluation Standards, 2nd Edition, 1994, and
- Generally Accepted Government Accounting Standards (GAGAS), widely referred to as the "yellow book" because of the distinctive color of its cover.

Other standards which are referenced in and adopted by GAGAS are:

- American Institute of Certified Public Accountants (AICPA) Statements on Auditing Standards, and Statements on Standards Attestation Engagements
- Codification of the Standards for the Professional Practice of Internal Auditing, promulgated by the Institute of Internal Auditors
- New Directions for Program Evaluation: Standards for Evaluation Practice, from the American Evaluation Association

Other evaluation standards are:

- Standards for Education Data Collection and Reporting (SEDCAR), produced by the Cooperative Education Data Collection and Reporting (CEDCAR) Standards Project Task Force, sponsored by the National Center for Education Statistics (NCES), U.S. Department of Education, September 1991
- Standards for Evaluations of Educational Programs, Projects, and Materials, developed by the Joint Committee on Standards for Educational Evaluation, 1981 (the precursor to the 1994 standards)
- The Personnel Evaluation Standards, developed by the Joint Committee on Standards for Educational Evaluation, 1988

DIALOGUE AND SELF-REFLECTION

Finally, and no less importantly, we engaged in continuing dialogue and self-reflection on the natures and differences of our fields. Much of this dialogue took the form of, "What we do is..." usually followed by, "We do that too, but..." Much of our dialogue, as mentioned above, sent us back to our fields' standards, but more profoundly, it caused us to reexamine our individual assumptions about the way in which our studies were conceptualized, how data should be collected and verified, how findings should be reported, and, overall, what our work meant to our organizations and to us personally. In the course of this dialogue and self-reflection, there was an ebb and flow in which one of us would conclude momentarily that her or his field had not gotten "it" entirely right, then would reverse course and find support for a strongly held, but not yet demonstrated, belief. By the end of the process, we believe we achieved a better rounded, more informed perspective about both fields which we hope is reflected in this paper.

RESULTS

Table 1 (attached) contains a comparison of evaluation and performance auditing with respect to their important features. Similarities and differences between the two fields are elaborated below.

DEFINITIONS

Table 1 provides the definition of auditing found in GAGAS. Many definitions of evaluation have been proffered. Several are reproduced in the table. It is difficult to offer a single, accepted definition of evaluation because what evaluation is depends to some extent on the model, the conceptual framework, from which it operates. The major audiences, assumptions, methodology, outcomes, and typical questions differ depending on the type of evaluation being conducted (House, 1980).

MODELS

A great deal of confusion attended our discussion of models. Evaluation has recourse to a large number of models, many developed by and associated with prominent evaluation practitioners (e.g., Provus' model, Stake's model, and Stufflebeam's CIPP model); however, the concept of "model" does not carry much weight in auditing. It is not, however, that auditing operates in the absence of a model, more that auditors do not think of their work in those terms. In fact, as a recent GAO publication points out:

> "An audit model--the "criterion, condition, cause and effect" model--is a special case of the criteria-referenced design that is widely used in GAO. Outcomes, the condition, are often compared to an objective, or a criterion, and the difference is taken as an indication of the extent to which the objective has been missed, achieved, or exceeded. However, it is not ordinarily possible to link the achievement of the objective to the program, because other factors not accounted for may enter into failure or success in meeting the objective" (GAO, 1991).

From its inception, ORE consciously selected a decision-making model. It is from that perspective that much of the succeeding analysis proceeds. However, as Sonnichson (1990) rightly observes in discussing the differences between evaluations and the "inspections" conducted by federal Inspectors General (OIG):

> "No claim of superiority for one version of evaluation over another is sustainable, and all such claims are clearly inappropriate. Creating a dialogue with OIG evaluators can contribute to a better understanding of their environment and to further development of the field of evaluation.
>
> "Evaluation, stripped of its jargon and fixation on methodological subtleties, essentially aims to produce information about how programs work and their effectiveness in alleviating the problems they are designed to cure. With this premise in mind, the evaluation profession ought to encourage any systematic review of programs that use evaluation to improve government services. No single evaluation approach can accomplish this goal, and all opportunities for professional development need to be examined for their potential contribution to this goal."

1988). Although the federal Inspectors General were created by the Inspector General Act of 1978 (Public Law 95-452) to (among other things) promote the economy, efficiency, and effectiveness of federal operations, many of their evaluation efforts focus on *cost saving* opportunities (as distinguished from *cost-effectiveness*.) (Cordray, 1990). One explanation for this different focus may be the proportion of accountants who were conducting financial and financial-type (GAGAS, 1994) audits prior to the IG Act. According to Sonnichson, "The nuclei of OIG staffs, in many cases, are investigators and auditors assigned to those functions prior to...1978...(Sonnichson, 1990). As Davis (1990) reports, "Without doubt, the disciplinary backgrounds of program evaluation staff affect their approaches to program evaluation." He also noted that auditors are more concerned with accountability, emphasizing management control rather than measuring program impact.

PURPOSES

As much as differences arise from the models they employ, evaluation and auditing differ even more fundamentally in how each profession expects its products to be used.

Impact

One notable difference between auditing and evaluation is in the effect they expect their work to have. Auditors expect the policies and practices of the organization to change as a result of their work. By the time the audit report is published, many of the deficiencies detected in the audit may already be remedied. While evaluators hope, not expect, that their work will be the basis both for program improvement and decisions about the continuance of programs, they do not want the experimental conditions to change during the course of the study. Program implementation may be improved during the course of an evaluation, especially in a program's formative stages, but if the experimental conditions change during the course of the study, the evaluation will be hampered in its summative purpose of making inferences about whether a program has caused the outcomes that have been observed. If causal inference is not possible, because the program changed too substantially to be taken into account by the evaluation, evaluators tend to regard their work as wasted.

Recommendations

Related to their differences in expected impact is a difference between audit and evaluation in the type of recommendations they make. Most evaluation reports contain recommendations, but they are more frequently of a more general, rather than a specific, nature. An evaluator is more likely to recommend that the organization pay more attention to a particular problem area than to propose a specific solution for the problem. Indeed, because many of the programs evaluators study are aimed at ameliorating broad social problems, it is reasonable to ask what sort of recommendations an evaluator could make.

Auditors, on the other hand, make specific recommendations for program improvement to which they expect a response from management. Management may concur or not (at the risk of unfavorable mention from policy makers), but auditors are expected to propose solutions to the problems they uncover.

OBJECTIVES

The five major objectives of auditing are enumerated in Table 1 and have already been mentioned. Evaluation shares with auditing the examination of what Chelimsky (1985) terms "normative" questions, i.e., the extent to which programs are in accord with established standards. However, where this is a compliance issue for an auditor, an evaluator is apt to be more concerned with establishing grounds for the failure of a program to accomplish its goals. The accomplishment of goals is also a shared objective, but evaluation concerns itself not only with whether a program accomplished what it was supposed to but also what it actually did accomplish, intended or not. Considerable program description is routine for an evaluation.

Evaluation departs from the audit objectives of safeguarding of assets and measuring economy and efficiency. Though efficiency might receive some passing mention in an evaluation, evaluators generally give the most attention to establishing the effectiveness of programs, though not as frequently to the extent of examining cost-effectiveness to see whether a program is worth the investment.

ASSUMPTIONS

Perhaps the most substantive difference between the professions arises from the assumptions they make, from the way in which auditors and evaluators regard their work. Evaluators, typically trained as social scientists, regard what they do as abiding in the broad realm of science and partaking of its methods, safeguards, and limitations. Auditors, on the other hand, regard themselves as serving as a tool of management, standing in relation to management as trusted advisors and counselors. "Auditor," in fact, means "listener."

This difference in orientation--academic/scientific versus service to management--underlies some notable contrasts between evaluation and performance auditing which are reflected both in their procedures and their products.

Evaluators do not, for example, believe they have the last word on a subject. Having conducted a study, they report their methods, their findings, and their interpretations, and invite critical review of their work by others in their field. They supply sufficient methodological detail for others to understand and, if desired, to *replicate* their work. By publishing their work, they intend it to contribute to the body of knowledge, not to constitute a definitive and final treatment of an area of study.

Auditors, by contrast, expect their work to be the definitive treatment of the subject they are auditing. Because they are licensed professionals whose work is periodically reviewed by their peers, auditors expect their work to be accepted on its face; they do not expect it to be examined by other auditors, nor to make a contribution to a general body of knowledge. Their work is validated by the standards of their profession and by the care with which it was carried out. Auditing is not concerned with the "replicability" of an experiment or its findings. Instead, auditors are asked to give their clients the last word, the proverbial bottom line. In short, auditing is not doing science; it is doing management consulting.

One consequence of this difference in orientation is that auditing does not tend to do "on the edge" analysis; it does not take a lot of risks. While an evaluation might concern itself with a broad area of social concern and conclude simply with a determination that a particular program or

intervention did not ameliorate the problem, without a specific recommendation of the means to solve the problem, auditing would not tend to engage a problem which was not susceptible to recommendations to management as to ways to address the problem. An evaluator might feel comfortable saying, in effect, "I do not know how to keep students from taking drugs, but this program did not have a significant effect on their drug-taking behavior," but an auditor would not. In auditing, it is not acceptable to send management "back to the drawing board." In this sense, the DARE audit was not a typical audit, or entirely satisfactory to the auditors, in that they were forced to make more general recommendations that they would have liked.

AUDIENCE

Both professions have multiple audiences, although they perceive the ultimate beneficiaries of their products differently. For evaluation, the audience is ultimately the "field" or the "literature." Program administrators, District administration, and the Board of Trustees (collectively, District "management") are intended recipients of evaluation information, but evaluators also perceive themselves to be contributing to an increase in knowledge about a specific area of educational research, evaluation, and practice. By contrast, because of its perception of itself as a tool of the policy makers for accountability purposes, auditing regards its principal audience as "management," to whom it makes recommendations for the remediation and improvement of deficiencies. The OCA views its primary audience as the City Council, and City management to be the user of the recommendations.

Reporting

To enable the general public, as well as colleagues in the research and evaluation community, to have access to their work, evaluators feel strongly about publishing and disseminating their findings through a variety of outlets. Evaluation reports written to be accessible to a general audience are widely disseminated, while technical reports documenting specific methodology are made available to the more limited audience of fellow practitioners, academics, and students to whom they are of interest. Many evaluation reports are disseminated through ERIC. Evaluators also often share their findings through presentations at professional conferences (e.g., AERA, AEA, and APA) and articles in professional journals.

Audit reports, on the other hand, are written and intended for management and policy makers. Reports are disseminated to a wide range of government administrators and policy makers, as well as the general public, but do not make their way into any sort of auditing "literature."

A number of similarities and differences relating to reporting are discussed under "Standards" below.

Authorship

An interesting sidelight to the difference in reporting style has to do with authorship. Because of their academic orientation, publication is the life's blood of evaluators, and they take professional pride in their publications. Indeed, publication is one indicator of an evaluator's productivity. Auditors do not claim authorship; their reports are the product of the audit team. While both audit and evaluation reports are published by their organizations, and in fact belong to them, evaluators regard their reports as their intellectual property, to be included in vitae and cited by the names of the authors. Auditors do not include their reports in their résumés, nor are reports cited except by

institutional author. (Indeed, audit reports do not typically contain a bibliography, nor do auditors expect their reports to be cited in other reports.)

In discussing how to cite the DARE audit report in the bibliography to this paper, something like the following conversation ensued:

HE: Who are the authors?

SHE: Authors?

HE: Who wrote the report?

SHE: We all wrote the report.

HE: Then who should be first author? Who wrote the most?

SHE: Well, in this instance, I suppose I wrote the most.

HE: Then you should be first author.

SHE: No, we don't do it that way. The report is the product of the office.

HE: Then I guess the office is the author of the report. Is that how you cite other audit reports?

SHE: Who would want to cite an audit report?

HE: Well, I would, but never mind. It just seems a heckuva way to run a railroad.

STANDARDS

Adherence to Standards

One significant general difference between the two professions is that audits conducted by the OCA must adhere to government auditing standards, while evaluation practitioners are merely urged to follow their field's standards. Indeed, one can be a working evaluator and be virtually unaware of the published standards. (At the 1994 AERA convention, at which the new standards were introduced, one experienced, long-time director of a well reputed school district R & E expressed surprise that he had never heard of the 1978 Joint standards.)

Not all governmental auditing shops adhere to the GAGAS; some follow the Standards for the Professional Practice of Internal Auditing promulgated by the Institute of Internal Auditors; others, particularly those whose functions are concentrated solely in the area of financial and financial-type auditing, stick with the AICPA's Statements on Auditing Standards. However, under the requirements of the Austin City Charter, the Office of the City Auditor must conduct its audits in accordance with GAGAS. ORE has long adhered to the 1978 Joint evaluation standards and began a formal adoption of the 1994 standards.

The authors' personal experiences with standards differ. In OCA, adherence to standards helps the auditors from making procedural errors and therefore keeps them out of trouble. A peer review performed every four years constitutes a compliance audit of the office to determine if procedures are carried out in accordance with the GAGAS. A "clean bill of health" amounts to an

endorsement of OCA's operation. In ORE, however, knowledge of the standards by nonevaluators has served simply as ammunition for critics to nit-pick. If a particular evaluation has not complied perfectly with any selected standard, regardless of the limitations placed on the evaluation by available resources, people who dislike evaluation can claim that the work is "flawed" and that the findings should be dismissed.

Sanctions for not following the standards also appear to be more severe for auditors than for evaluators. While criticism may sometimes be directed at evaluation studies based on a perceived failure to comply fully with a specific evaluation standard, significant noncompliance with GAGAS can place the survival of the entire OCA, or at least the careers of its executive management, in jeopardy. Moreover, individual auditors who have obtained either the CPA or CIA certification may lose that certification if their professional conduct is found to be in violation of their respective organization's standards or codes of conduct. The profession expects the Certified Internal Auditor to abide by the:

- Internal Auditor Code of Ethics,
- Statement of Responsibilities of Internal Auditing, and the
- Standards for the Professional Practice of Internal Auditing.

As for governmental auditing offices, the Yellow Book calls for an external quality control review to be conducted every three years on audit organizations which perform audits in accordance with GAGAS. The purpose of the external review is to "determine whether the organization's internal quality control system is in place and operating effectively to provide reasonable assurance that established policies and procedures and *applicable auditing standards* are being followed" [emphasis ours] (GAO, 1994).

Comparison of Standards

A considerable number of similarities exist between the Program Evaluation Standards and Governmental Auditing Standards, but as with other aspects of these two fields, the differences are remarkable and enlightening. Similarities exist in requirements to identify and communicate thoroughly with stakeholders during the course of designing an evaluation or a performance audit; and to ensure the reliability, validity, and relevance of the information used in forming judgments or conclusions. Moreover, both call for fairness in reporting strengths and weaknesses of the subject under study; and both require that reports be clear and issued in a timely manner to enable use of the results.

However, in addition to the similarities between the two, some contrasts come immediately to mind. The Yellow Book includes considerable guidance on the operation of the audit office, including requirements that the: qualifications of staff assigned include "adequate professional proficiency for the task required"; relationship of the audit office to its parent organization be structured to ensure adequate organizational independence; and that an internal quality control system be in place to ensure compliance with applicable standards, and to provide reasonable assurance that "adequate audit policies and procedures have been established and are being followed." By contrast, the Program Evaluation (PE) standards do address competence of staff and fiscal responsibility in expenditure of resources, but no other aspects of managing an evaluation organizaiton.

Selected standards from the Yellow Book (1994 revision) and the Program Evaluation Standards (2nd edition) are cited below, together with discussion of their compared characteristics.

Ensuring the reliability and validity of evidence used to draw conclusions. While Yellow Book standards outline certain general presumptions to aid in judging the competence (validity and reliability) of evidence, they appear to rely more heavily on the auditor's judgment in developing data collection methods and ensuring the adequacy of evidence for drawing conclusions. One exception to this reliance on auditor judgment in the Yellow Book is an extensive discussion on the need for ensuring the reliability of evidence from computer-based systems.

GAGAS: Field work standards: 6.46. Evidence
Sufficient, competent, and relevant evidence is to be obtained to afford a reasonable basis for the auditors' findings and conclusions. A record of the auditors' work should be retained in the form of working papers. Working papers should contain sufficient information to enable an experienced auditor having no previous connection with the audit to ascertain from them the evidence that supports the auditors' significant conclusions and judgments.

Subsequent discussion and explanatory guidance on evidence in GAGAS comprises a total of 26 paragraphs, of which 13 have two sentences or less. This relatively short text contrasts with the PE standards' 45 paragraphs (many very lengthy) which discuss guidelines and common errors in assuring validity and reliability –and this count does not include the illustrative cases in both chapters. Table 2 summarizes the guidance provided in the Yellow Book related to evidence (other than that addressing reliability of computer-generated data).

TABLE 2
Summary of guidance provided in GAGAS on types of evidence and tests of evidence (except evidence obtained from automated systems)

CATEGORIES OF EVIDENCE:	TESTS OF EVIDENCE:
Physical	Sufficient (enough to support findings)
Documentary	Relevant (logical relationship to finding)
Testimonial	Competent (consistent with fact–valid, reliable)
Analytical	

SOURCE: Author analysis of *Generally Accepted Governmental Auditing Standards*, GAO, 1994 Revision

The additional discussion under the GAGAS Evidence standard provides auditors with guidance on how to judge the *competence* of each of the types of evidence. For example,

> "Evidence obtained from a credible third party is more competent than that secured from the auditee..."
> "Evidence obtained under an effective system of management controls is more competent than that obtained where such controls are weak or nonexistent..."

OCA auditors seldom have the luxury of obtaining evidence under an effective system of management controls, since the audit planning practices of the Office have been to maximize the use of limited audit resources by focusing on high-vulnerability systems. However, with respect to establishing certain elements of an audit finding, the use of outside information is common to almost all audits.

For instance, to develop *criteria* ("how it should be") OCA auditors frequently look to performance information available from other governmental entities. Because "industry standards" and identified "best practices" often do not provide data against which to compare the efficiency and effectiveness of certain types of *government* operations, comparisons to a variety of similar activities carried out in other jurisdictions frequently constitute the best available data for comparing with local programs. However, unless other audits or evaluations exist which have designated a particular government program as the benchmark, these comparisons may provide little more than a starting point for auditors, by revealing the types of questions that will need to be asked to determine causes for specific variances.

In addition to use of external information for establishment of criteria, conclusions developed from use of extra-organizational data for evidence of *condition* ("how it really is") and sometimes *effect*, ("what it leads to") are also considered stronger. The DARE audit, for example, relied on juvenile offense data from the Travis County Juvenile Court, and data from the District's student records data base for information on academic performance and disciplines to assess program effectiveness. These data were used to determine program effect–to answer the question, "Has exposure to DARE reduced juveniles' contact with Travis County Juvenile Court, or improved selected academic behaviors, compared to students who were not exposed to DARE?"

Prior evaluations of Austin's DARE program, carried out by ORE, were also cited in the audit report, and incorporated into the conclusions. Although the evaluations (which included annual snapshots of academic and discipline performance from the same District data base used in the OCA DARE audit, plus results of opinion and self-reported behavior surveys of exiting DARE students) did not employ the experimental design utilized by the auditors, the results were *relevant* and thus, in accordance with standards, needed to be considered in the auditors' overall formulation of conclusions.

Besides the discussion of competence based on sources, the Yellow Book provides certain "tests" of the sufficiency, competence, and relevance of data based on the methods used to gather it. For example:

- *Data gathered by auditors:* (through questionnaires, structured interviews, observations, and computations) "The design of these methods and the skill of the auditors applying them are the keys to ensuring that these data constitute sufficient, competent, and relevant evidence."
- *Data gathered by the auditee*: "Auditors may determine the validity and reliability by direct tests of the data...When the auditors' tests of data disclose errors, they may find it necessary to:...seek evidence from other sources,...redefine the audit's objectives to eliminate the need to use the data, or...use the data, but clearly indicate in their report the data's limitations and refrain from making unwarranted conclusions or recommendations."
- *Data from computer-based systems*: "Auditors should obtain sufficient, competent, and relevant evidence that computer-processed data are valid and reliable when these data are significant to the auditors' findings."

In auditing, one of the most commonly used "direct tests of the data" to determine validity and reliability is *reconciliation.* Auditors will frequently reconcile reported totals to the component amounts, or compare amounts reported in one document to similar or identical data used in other documents. This practice can frequently result in considerable efforts when variances are found. The efforts will be expended by both auditees and auditors to identify the sources of specific data sets and disaggregate their constituent elements to assess the validity of a particular field in reporting on the activities or results being evaluated.

Examining the competence of data frequently involves the spreadsheet, another test of data frequently used by auditors to determine the variances between computerized or other reports and their original documents. An unusual spreadsheet was used in the DARE audit to determine the extent of the variance between demographic characteristics of two populations. Specifically, the experimental design of the audit required that both populations used in the comparison be demographically similar. (These two populations were the DARE students who were fifth graders when DARE was piloted in 1987-88–the experimental group–and the Non-DARE–or control group students–who were fifth graders in 1986-87, the year before DARE was piloted). Because the auditors were comparing students from the same schools only one year apart, they initially assumed that the populations would be similar, because they came from the same attendance area. However, the two groups varied demographically.

The reasons for the demographic differences could not be identified until late in the audit, when one participant expressed concern about the effects of *busing* on the comparability of the two populations. Pressing further, the auditors learned that busing had been carried out until 1986-87 (the control group year), and eliminated in the District in 1987-88 (the experimental group year). The effect of busing was to modify the demographic makeup of several individual schools, a condition which could be the cause of the observed variances between the two cohorts being compared.

Additional efforts to assess the competence and relevance of the evidence were needed, and first, the documentation of the District's busing plan was reviewed in detail, with the auditors constructing a rough spreadsheet of the demographics of each school before and after busing, to analyze the changes that had occurred at each school when busing was eliminated. To their relief, the auditors learned that only five of the 15 schools in the audit had been impacted by the switch from busing to neighborhood schools. Moreover, the auditors were able to remove one school from the audit and obtain an acceptable degree of difference between each total cohort (DARE and Non-DARE). The two cohorts' overall similarity was achieved with relatively little reduction of the sample size because the District's busing approach in 1986-87 and before had used "paired" schools. In other words, two separate school attendance areas–one containing a large minority population, and one with a small minority population–were paired for exchange of students during busing. Coincidentally, of a total of 15 DARE pilot schools in the audit, only five had been included in the District's busing plan. Luckily, four of the five schools had included both of the "paired" neighborhoods. The auditors eliminated the one school whose paired school had not piloted DARE, and the demographics of the total cohorts became statistically similar. With the odd school removed, demographic comparisons of the two (total) cohorts revealed that the change from busing to neighborhood schools had had little impact on the minority make up (60.3% DARE, 61.6% Non-DARE) and on the low-income status (28.3% DARE, 24% Non-DARE) of the total cohorts.

Other tests of evidence typically carried out by auditors include various types of ratio analyses. For instance, to test the reasonability of an annual revenue figure, auditors might calculate the average daily revenue, or the average revenue per customers served.

The PE standards do not appear to place the same reliance on the judgment of the evaluator as is found in the Yellow Book. Instead, the process of validating data is described in detail, along with extensive guidelines to assist evaluators in ensuring validity and reliability. The standard which calls for systematic review and elimination of errors instructs evaluators in the approaches and necessity for scrubbing data as thoroughly as possible.

PE Standards, Accuracy: A5. Valid Information
The information gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use.

PE Standards, Accuracy: A6. Reliable Information
The information gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use.

PE Standards, Accuracy: A7. Systematic Information
The information collected, processed, and reported in an evaluation should be systematically reviewed and any errors found should be corrected.

Objectivity in reporting To their credit, both sets of standards emphasize the necessity for objective reporting, first by requiring practitioners to describe the methodology and any related limitations (either of methods or of data). In addition, both sets of standards prescribe a fair discussion of the conclusions which clearly places the extent of any problems identified in the context of the overall conditions observed.

GAGAS Reporting Standards for Performance Audits: 7.10 Objectives, Scope, and Methodology
Auditors should report the audit objectives and the audit scope and methodology.

Supplemental guidelines addressing scope require auditors to "describe the depth and coverage of work...the relationship between the universe and what was audited...report the kinds and sources of evidence...and explain any quality or other problems with the evidence. Auditors should also report significant constraints imposed on the audit approach by data limitations or scope impairments."

Further text which discusses the report of methodology requires auditors to "clearly explain the evidence gathering and analysis techniques used. ...identify any significant assumptions made in conducting the audit...and when sampling significantly supports auditors' findings, describe the sample design and state why it was chosen."

The GAGAS standard for reporting objective, scope, and methodology, together with the guidance that supports it appears to be substantially congruent with similar requirements in the PE Standards:

PE Standards: Accuracy : A4. Defensible Information Sources
The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.

The GAGAS standard and supplemental guidance on reporting of audit results appear to cover much the same territory as comparable PE standards. However, where PE standards place responsibility on the evaluator to report results objectively, the auditors must take one step further. To wit, the "views of responsible officials" (the auditee management's response to the report) must be included in the report.

GAGAS Reporting Standards: 7.17 Audit Results
Auditors should report significant audit findings, and where applicable, auditors' conclusions.

The supplemental text which expands on the Audit Results standard requires auditors to "include sufficient, competent, and relevant information to promote adequate understanding of the matters reported and to provide convincing but fair presentations in proper perspective..." This section of the Yellow Book also describes the elements of an audit finding ("criteria, condition, and effect, plus cause when problems are found...the elements needed for a finding depend entirely on the objectives of the audit.") and discuss the reporting of conclusions. "Conclusions are logical inferences about the program based on the auditors' findings. Conclusions should be specified and not left to be inferred by readers. The strength of the auditors' conclusions depends on the persuasiveness of the evidence supporting the findings and the convincingness of the logic used to formulate the conclusions."

In addition to the auditor's requirement to be objective and convincing implied in the descriptive text under the Audit Results standard quoted above, it is part of a standard on report presentation, as well.

7.50 Report Presentation
The report should be complete, accurate, objective, convincing, and as clear and concise as the subject permits.
The 15 paragraphs that follow this standard call for "perspective on the extent and significance of reported findings...correct portrayal of findings...If data are significant to the audit findings and conclusions, but are not audited, the auditors should clearly indicate in their report the data's limitations and not make unwarranted conclusions or recommendations based on those data."

However, the intent of the Yellow Book to ensure objective reporting is most forcefully brought home in its requirement to include auditee responses within the audit report itself.

GAGAS Reporting Standards: 7.38 Views of Responsible Officials
Auditors should report the views of responsible officials of the audited program concerning auditors' findings, conclusions, and recommendations, as well as corrections planned.

The Yellow Book explains: "One of the most effective ways to ensure that a report is fair, complete, and objective is to obtain advance review and comments by responsible auditee officials and others...When the comments oppose the report's findings, conclusions, or recommendations, and are not, in the auditors' opinion, valid, the auditors may choose to state their reasons for rejecting them. Conversely, the auditors should modify their report if they find the comments

valid." The DARE auditor in charge has come to describe this particular requirement as an "insurance clause," because of the numerous times it has prevented errors from being published in the final report. One incident, early in her career, has indelibly demonstrated the value of this step...

The occasion when erroneous audit conclusions were publicized resulted from a communication breakdown between the auditee's official liaison to the audit and the organization being audited. The liaison assigned to answer auditors' questions and assist them in locating data needed for an audit of a state agency's fleet operations had recently (unbeknownst to the auditors) been demoted and offered early retirement. This person, as his last assignment before retiring, was responsible for explaining agency fleet management practices to the audit team. Among the findings auditors reported was the damning revelation that the agency did not retain documentation of fleet maintenance activity for use in monitoring cost-effectiveness of maintenance investments.

When the auditors' oversight body requested an interim briefing on audit findings, the auditors had a draft report in final stages, almost ready to provide to management for response. This draft they provided to the legislative committee, although all copies were numbered and requested to be returned at the conclusion of the briefing. However, one copy was not returned, and the audit director decided not to attempt to retrieve it. That copy had been given to a guest of the legislative committee–a House member–who had been highly critical of the auditee agency for quite some time.

The next morning the headlines read: "[Agency] Director Denies Audit Allegations." The legislator had leaked the draft report to the press, who had then contacted the agency head for his comments. The director vigorously denied the audit's finding that vehicle maintenance documentation was not retained, along with a host of other audit findings that he claimed were "misrepresented." At a hastily-called meeting between auditors and the agency staff, the auditors learned that their liaison had misled them, deliberately obfuscating relevant information in order to embarrass his bosses. The damage that resulted so severely crippled the audit department's credibility that the entire function was eliminated in the very next budget session.

This lesson was learned early in the auditor's career, and never forgotten. Findings and conclusions of every audit she works on are discussed early in the process, continuously throughout the process, and repeatedly, with everyone in the auditee organization that will sit still for a briefing. "*No surprises*" is one of the main mottoes of the OCA. (The other principal motto of the office is "*trust, but verify.*")

The Yellow Book calls for a balanced tone in reporting, reminding auditors to "keep in mind that one of their objectives is to persuade, and that this can best be done by avoiding language that generates defensiveness and opposition." The emphasis on objectivity in GAGAS is paralleled in the PE standards which address reporting on conclusions:

PE Standards, Accuracy: A10. Justified Conclusions
The conclusions reached in an evaluation should be explicitly justified, so that the stakeholders can assess them.

PE Standards, Accuracy: A11. Impartial Reporting
Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.

PE Standards, Utility: U4. Values Identification
The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.

PE Standards, Utility: U5. Report Clarity
Evaluation reports should clearly describe the program being evaluated, including its context, the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.

Recommendations

With respect to the *making* of recommendations, while auditors' standards differ from those of evaluators, when it comes to *seeing recommendations implemented*, actual experiences of both professions may be more congruent than their standards would suggest. Auditing standards require auditors to make recommendations when opportunities for improvement are identified, while PE standards merely reference the "likelihood that the evaluation will be used..."

GAGAS, Reporting Standards for Government Audits: 7.21 Recommendations
Auditors should report recommendations for actions to correct problem areas and to improve operations.

PE Standards, Utility: U7. Evaluation Impact
Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased.

Most auditors, when they meet, will find some opportunity to decry the duplicity of "management" [the auditee], who has learned to agree with the recommendations in writing, so that when responses are published in the final audit report, management's plan of action is clearly designed to fully resolve the problems identified. Yet, when auditors subsequently follow up on actual implementation, nothing has been done. This is the auditor's lot–to find ways to make government run for efficiently, economically, and effectively, only to see "business as usual" continue unimpeded.

In the last two years, the OCA has, like Popeye, eaten a can of "spinach" and has suddenly thrown off the traditional impotence that burdens the auditor's life. This new power has arisen from two fairly recent occurrences. First, in a 1991 plebiscite, the citizens of Austin approved a revision to the City Charter that caused the City Auditor to be appointed by and begin reporting directly to the City Council. Previously, the City Auditor was appointed by the City Manager. This change in itself had a significant impact on the relationship between management and auditing. (Not surprisingly, that relationship has become somewhat more adversarial.) Then, in 1992, the City Council approved a resolution which requires City Management to report semiannually on its progress in implementing audit recommendations, a report which must be verified by the City Auditor. Notwithstanding the semiannual requirement of the resolution, the current City Council

Audit and Finance Committee has instructed the City Auditor to report quarterly on its follow-up of audit recommendations.

The increase in management accountability which resulted from these two developments has profoundly changed the reception given to audit recommendations by City officials. Prior to the 1991 election, management practiced the age-old tradition of agreeing with audit recommendations but failing to follow up with actual implementation. Since management is now being held increasingly accountable for either accomplishing its agreed-upon corrective measures or explaining to the City Council why it was unable to do so, the attention paid to the recommendations during the report drafting stage of the audit has heightened significantly. (To gain an accurate perspective on the relationships within City government, it is necessary to understand that Austin's City Council has adopted a mistrustful attitude towards management, occasionally requesting the audit staff to verify information provided by the City Manager's Office.)

The clout currently being experienced by Austin's City Auditor's Office is not shared by all local government audit offices. Auditor reporting relationships, and the degree of accountability on management for following up on audit recommendations vary considerably among jurisdictions. Needless to say, the ORE staff who worked with the auditors on the DARE and Partnered audits expressed a degree of envy for the City Auditor's enforcement muscle.

Providing Assistance to Management Subsequent to the Evaluation or Audit

Unlike the PE standards, the Yellow Book does not provide direction on the issue of assisting management in implementing audit recommendations, although it does raise questions about the ability of auditors to be independent if they are auditing a program or operation over which they previously had management responsibility.

In the auditing world, there are two opposing viewpoints on the subject of assistance to management. In this debate, objectivity and independence are pitted against the increasing pressure to take a more proactive role in the larger organization. Currently, a raging debate is under way in the auditing profession on the subject of assistance to management. As previously discussed, one principal aspect of the continuing struggle within the profession to chart its course in the current era of government and corporate downsizing is a dawning recognition of the imperative need for an audit function to "add value" to its organization in order to secure its survival. Among the arguments about how to add value are those which contend that the auditor's extensive knowledge of effective management practices and performance measurement techniques places an obligation (or at least uniquely positions the auditor) to assist management, in a role often called "internal management consulting." (McNamee, 1993; Ratliff, 1994; Makosz, 1994)

GAGAS, General Standards: 3.11 Independence
In all matters relating to the audit work, the audit organization and the individual auditors, whether government or public, should be free from personal and external impairments to independence, should be organizationally independent, and should maintain an independent attitude and appearance.

Among the types of personal impairments described in the supplemental text, GAGAS considers one type of impairment to be: "previous responsibility for decision-making or managing an entity that would affect current operations of the entity or program being audited. The PE standard

which directs evaluators along similar lines is U7, Evaluation Impact (quoted above in the section on recommendations). The text of the standard itself does not directly address independence, but the overview text admonishes evaluators not to "take on the role of the client..." Moreover, under the section describing "Common Errors" relating to this utility standard, is included "Taking over the client's responsibilities for acting on evaluation findings."

ARENA OF OPERATION

Audit operates both in the business and governmental arenas. Evaluation operates primarily in the government arena, although persons with evaluation skills are sometimes employed by business to conduct market research.

AUTHORITY

Two aspects of authority are worthy of discussion: (1) the warrant by which an audit or evaluation unit is authorized to operate, and (2) the perceived authoritativeness of its findings. The authority to conduct an audit or evaluation derives from the unit's parent organization which defines the unit's scope of operations. ORE and OCA are similar in that they are both internal to the parent organizations. ORE reports to the CEO of the organization, the CA is hired by and serves at the pleasure of the City Council, the "Board of directors," or the elected officials which set policy for the City organization. However, where ORE's mission was self-defined and is supported only by an administrative regulation and custom, OCA has a comprehensive charter from the City. The authority experienced by Austin's City Auditor's Office is not shared by all local government audit offices, however. Auditor reporting relationships, and the degree of accountability for follow-up vary considerably among jurisdictions.

The authoritativeness of both evaluation and audit findings derive from the perceived competence of its practitioners and from the credibility of their reports. However, in the case of auditors, their authority is buttressed by the licensure of practitioners, as well as the annual requirement for professional development and periodic peer review. ORE does have a oversight body, the Evaluation Advisory Committee, which is composed of educators, parents, and concerned citizens, but it functions in an advisory capacity and cannot certify professional competence.

TRAINING OF PRACTITIONERS

As previously noted, despite the growing pressures that are forcing changes in the scope and objectives of audits, most internal auditors have an accounting or auditing background. Many have the Certified Public Accountant (CPA) designation, and/or a Certified Internal Auditor (CIA) designation. Obtaining a CPA requires a greater than ordinary interest in accounting. To qualify for a CPA license in most states, candidates must pass the CPA examination and work for a minimum of two years in a related function, supervised by a CPA. But before a candidate can sit for the CPA examination, he or she must have successfully completed a minimum of 32 hours of college-level accounting courses.

GAGAS requires all persons who conduct government audits to complete a minimum of 80 hours of continuing professional education every two years, with at least 20 of those hours occurring in any single year. CIAs are also required by their licensing organization to maintain continuing professional education, and report a minimum of 100 hours every two years.

Most evaluators have a social sciences background, with advanced degrees (a masters at the minimum) in psychology, sociology, educational psychology, and related fields. In hiring evaluators, a premium is placed on graduate-level course work in research design, statistics, and quantitative methods. Evaluators are not certified by a professional organization, nor do they have a requirement for continuing professional education.

CONCLUSIONS

From our collaboration and study of performance auditing and evaluation, we have concluded that:

Evaluation and performance auditing, in their best expressions, are very much alike.

- Both fields are rule governed.
- Both are systematic investigations.
- Both are concerned with rules of evidence--its sufficiency, relevance, and validity.
- Both carefully document the steps in their work.
- Both compare actual performance against a standard.
- Most important, both assess program effectiveness.

As similar as evaluation and performance auditing are, there are , nonetheless, notable differences between them in philosophy and methodology.

The prime difference is that auditing cannot address cause-and-effect questions. As Chelimsky (1985) cogently explains:

"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...." (p. 489)

The professions could both profit from adopting methodology from one another.

Evaluation does not engage in formal analysis of risk. Use of this methodology would enhance the value of evaluation to its parent organization by providing a systematic means to allocate resources. Even an organization like ORE, which annually sets out a comprehensive agenda of its work for review by administrators and the Board of Trustees, could ultimately justify its allocation of resources only on the basis of perceived importance, not as the result of an empirical analysis of benefit to the District.

Auditing could learn more about the formal testing of instruments for reliability and validity. Because auditors are concerned about the meaningfulness of evidence, they need to understand how the quality of instruments relates directly to the quality of the conclusions which can be drawn from the information collected.

Evaluation could learn from auditing the extensive procedures it employs for ensuring the reliability of evidence from computer-based systems. While evaluators certainly check the reliability of data sources, and are concerned with the amount of error in measurement, they place much less emphasis on verification than do auditors.

Evaluation should consider adopting as a standard procedure reporting which includes management responses. Although some evaluation units do incorporate responses from program administrators in their reports, and some even have built in a cycle of management response to their findings, evaluation findings are more frequently met with no or an uncoordinated response. Though evaluators must guard against "taking over the client's responsibilities for acting on evaluation findings," they must nonetheless seek to have their findings utilized, else the entire enterprise is worthless.

Evaluation exists in a greater variety of forms (models) than auditing. Although auditing does not operate in the absence of a model, as previously noted, it does not as self-consciously direct its operations with reference to models as does evaluation. In this, it seems to us, auditing lacks a certain conceptual breadth, though, conversely, it may enjoy more acceptance by users who know what they are getting.

In the study of a program, the different viewpoints brought by professionals from different organizations are useful.

Closer cooperation between school district R & E offices and district auditors might serve to leverage their resources and maximize their information products. While it is nowhere near as simple as having the auditors "look at the money" and the evaluators at effect, and then combining them, auditing and evaluation methodologies could both be utilized to produce a greater depth of information than either might alone.

DISCUSSION

Evaluation has been occurring in most major American school districts since the 1970's. Early, model-driven attempts at evaluation have been replaced by more eclectic approaches. At the same time, performance auditing is coming to replace traditional financial auditing as the most prevalent form of auditing. One indicator of this trend is the increase in the number of evaluators at GAO since 1980. Yet, the practitioners of these professions seem to know surprisingly little about one another's fields, even though they may be studying the same kinds of programs, sometimes with similar methods.

As governmental bodies, such as school boards and city councils, strive to make decisions about the effectiveness of expensive programs and initiatives, systematic, data-based inquiries, whether called evaluations or performance audits, are becoming more and more important. Understanding the commonalities and differences of audits and evaluation studies will help practitioners be more effective and will enable collaborative studies between governmental agencies which will provide more information more efficiently.

Table 1 Comparison of Auditing to Evaluation

	GOVERNMENT AUDITING		EVALUATION
Comparison of Evaluation vs Auditing			
	GOVERNMENT AUDITING		EVALUATION
DEFINITIONS	"All audits begin with objectives [which] determine the type of audit to be conducted.... The types of government audits, as defined by their objectives, are ... financial audits or performance audits."		Assessing the worth of government actions
	(GAGAS 1994 Revision)		(Davis, 1990)
	Internal Auditing is an independent appraisal activity within an organization for the review of operations as a service to management. It is a managerial control which functions by measuring and evaluating the effectiveness of other controls.		"The systematic application of social (science) research procedures in assessing the conceptualization and design, implementation, and utility of social intervention programs" (Rossi & Freeman, 1985)
	IIA Statement of Responsibilities		
	Webster: "1. An examination of records or accounts to check their accuracy. 2. To examine, verify, or correct. (fr. Latin auditus, a hearing)		"The systematic investigation of the worth or merit of an object."
	(Webster,1981)		(Program Evaluation Standards, 2nd Edition, 1994)
MODELS	Criterion, Condition, Cause, Effect		Various
PURPOSES	To ensure accountability of the executive functions for full and proper implementation of policy and stewardship for public assets.		To provide empirical information that is useful to policy makers and program managers.
			(Chelimsky, 1985)
	To identify improvement opportunities.		"Program evaluation alone...can address cause-and-effect questions."
			(Chelimsky, 1985)

Table 1 Comparison of Auditing to Evaluation

	GOVERNMENT AUDITING		EVALUATION
OBJECTIVES	**Safeguarding of Assets**		**Program description**
	Determine the extent to which controls are in place, functioning, and adequate to protect against loss of assets through various means, such as fraud, waste, or abuse.		(Chelimsky, 1985)
	Compliance		**Information for program improvement**
	Determine the extent to which programs or activities comply with established criteria, such as laws, regulations, or contract terms.		(Chelimsky, 1985)
	Accomplishment of goals and objectivies		**Program outcomes**
	Determine whether programs or activities are achieving established or intrinsic goals and objectives.		To determine all actual outcomes of a program, whether intended or not by the program's formal objectives
	Reliability and Integrity of financial or performance information		(Chelimsky, 1985)
	Whether performance information reliably reflects actual performance of the entity		
	Efficiency and Economy		**Attempt to determine cause and effect**
	Whether the entity is using its resources economically and efficiently.		(Chelimsky, 1985)
ASSUMPTIONS	Service to management/policy makers		Academic/Scientific
AUDIENCE	• Policy makers		• Policy makers
	• Agency/Program managers		• Agency/Program managers
	• Other auditors		• Research Community
	• General Public		• General Public
	• Special Interest Groups		• Special Interest Groups

Table 1 Comparison of Auditing to Evaluation

	GOVERNMENT AUDITING	EVALUATION
STANDARDS	• Generally Accepted Government Auditing Standards (GAGAS) and • Institute of Internal Auditors (IIA) Standards	• ERS/AEA standards; • Joint standards (1978 and 1994 revised) • SEDCAR standards
	Other standards adopted by reference:	
	Statements on Auditing Standards and Statements on Standards for Attestation Engagements (AICPA)	
	Standards for the Professional Practice of Internal Auditing (IIA)	
	Standards for Evaluation Practice (AEA)	
ARENA OF OPERATION	All activities, laws, policies, regulations, units, contractors, and agents of government.	All activities, laws, policies, regulations, units, contractors, and agents of government.
	Internal audit in private industry.	Use of evaluation in private industry for market research.
AUTHORITY	Granted by the auditor's contract (if an external, consultant entity) or by charter of the governmental entity in which the audit unit exists.	Granted by contract or organizational mission of evaluation entity
	Granted by the governmental entity's contract with the auditee entity.	
TRAINING OF PRACTITIONERS	Primarily Accounting and Auditing	Social Sciences
Procedures	• Survey • Fieldwork • Reporting	• Evaluation Design • Data Collection • Data Analysis • Reporting

BIBLIOGRAPHY

Alwin, L. F. (1992, February). Reinventing the governmental auditor. The Internal Auditor, 49 (1), 16-17.

Burnaby, P., Powell, N., & Strickland, S. (1992, February). Contrasts: Internal auditing in the U.S. and Japan. The Internal Auditor, 49 (1), 38-45.

Burns, S. J. (1991, April). Auditing high technology ventures. The Internal Auditor, 48, (2), 56-59.

Chelimsky, E. (1985, August). Comparing and contrasting auditing and evaluation: Some notes on their relationship. Educational Review, 49 (4), 483-503.

Chelimsky, E. (1988, Summer). Federal evaluation in a legislative environment: Producing on a faster track. In C. G. Wye & H. P. Hatry (Eds.), New directions for program evaluation: No. 38. Timely, low-cost evaluation in the public sector. San Francisco: Jossey-Bass.

Cordray, D. S. (1990, Winter). Inspections as program evaluation: A partial overlap. In M. Hendricks, M. Mangano, & W. C. Moran (Eds.), New directions for program evaluation: No. 48. Inspectors general: A new force in evaluation. San Francisco: Jossey-Bass.

Davis, D. F. (1990). Do you want a performance audit or a program evaluation? Public Administration Review, 50 (1), 35-41.

Flaherty, J. J., & Stein, J. (1991, October). Solution-focused audit reporting. The Internal Auditor, 48 (5), 58-61.

General Accounting Office. (1994, June). Government auditing standards 1994 revision (GAO/OCG-94-4). Washington, D.C.: U.S. Government Printing Office.

General Accounting Office. (1991). Designing evaluations. Washington, D.C.: U.S. Government Printing Office.

House, E. R. (1980). Evaluating with validity. Beverly Hills, CA: Sage Publications, Inc.

House, E. R. (1993). Professional Evaluation. Newbury Park, CA: Sage Publications, Inc.

Joint Committee on Standards for Educational Evaluation. (1994). The program evaluation standards, 2nd edition: How to assess evaluations of educational programs. Thousand Oaks, CA: Sage Publications, Inc.

Julien, F. W. (1993, August). The power of benchmarking. The Internal Auditor, 50 (4), 23-28.

Kusel, J., Oxner, T. H., & Rivers, R. (1992, June). The 1992 job market. The Internal Auditor, 49 (3), 21-30.

Lynch, J. J. (1992, April). Eliminate the auditors? The Internal Auditor, 49 (2), 27-32.

Makosz, P. (1994, August). The future of internal auditing. The Internal Auditor, 51 (4),

Malan, R. M. (1991, June). Internal auditing in government. The Internal Auditor, 48 (3), 90-95.

Malan, R. M. (1988). Local government evaluation in a legislative environment. In C. G. Wye & H. P. Hatry (Eds.), New directions for program evaluation: No. 38. Timely, low-cost evaluation in the public sector. San Francisco: Jossey-Bass.

McNamee, D. (1993, October). Calling the tough shots: Interview with Charles A. Bowsher. The Internal Auditor, 50 (5), 30-34.

McNamee, D., & McNamee, T. (1992, December). Breakpoint. The Internal Auditor, 49 (6), 26-30.

Ng Lau, N. (1994, August). Viewpoints: Accountability. The Internal Auditor, 51 (4), 23.

Office of the City Auditor. (1994). Opportunities for Youth II: Austin Police Department's Drug Abuse Resistance Education (DARE) program (Audit No. S9305D). Austin, TX: City of Austin.

Oxner, T. H., & Kusel, J. (1994, June). The 1994 job market, How green is your valley? The Internal Auditor, 51 (3), 24-34.

Paape, L. (1993, December). Seizing opportunities. The Internal Auditor, 50 (6), 51-56.

Parle, W. M., Wallace, B. C., & Davis, J. A. (1990, Spring). Auditor selection practices and use of expanded-scope audits by large American cities. State and Local Government Review, 55-58.

Ratliff, R. L., & Beckstead, S. M. (1994, December). How world-class management is changing internal auditing. The Internal Auditor, 51 (6), 38-42.

Sivaraman, S. M. (1992, June). Moonrakers or Sun raisers? The Internal Auditor, 49 (3), 43-47.

Sonnichsen, R. C. (1990, Winter). Another view of program inspections by the Offices of Inspectors General. In M. Hendricks, M. Mangano, & W. C. Moran (Eds.), New directions for program evaluation: No. 48. Inspectors general: A new force in evaluation. San Francisco: Jossey-Bass.

Stern, G. M. (1994, April). Fifteen ways auditing departments are adding value. The Internal Auditor, 51 (2), 30-33.

Stipak, B., & O'Toole, D. E. (1990, Spring). Performance auditing in local government: Current use and future prospects. State and Local Government Review, 51-54.

Thompson, S. A. (1991, April). Saving money, making money. The Internal Auditor, 48 (2), 45-49.

Wernz, J. (1994, August). Viewpoints: Change. The Internal Auditor, 51 (4), 19-20.