DOCUMENT RESUME

ED 251 509 TM 850 010

AUTHOR Herman, Joan; And Others
TITLE Whither Educational Evaluation: Life, Death, and Reincarnation? Research into Practice Project.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Apr 84
GRANT NIE-G-84-0112-P4
NOTE 58p.; Papers presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Collected Works - Serials (022)
JOURNAL CIT CSE Evaluation Comment; v7 n1 1984

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Educational Assessment; Elementary Secondary Education; *Evaluation Needs; Evaluation Utilization; *Evaluators; *Federal Government; *Financial Support; Higher Education; Instructional Improvement; Professional Recognition; Professional Training; State Government
IDENTIFIERS Evaluation Problems; Evaluation Research

ABSTRACT

The current health and future well-being of educational evaluation is considered in terms of financial support, official mandate, evaluators' status, evaluation prevalence, and evaluation utility in these seven symposium papers. "Educational Evaluation--Indicators of Life, Death, or Reincarnation," by Susan Klein, assesses recent federal and national trends. "Is There a Perceived or Real Metamorphosis of Purpose for Educational Evaluation?" by Eva L. Baker, sees the need to reconcile both broad policy issues and local utilization needs. "What is the Prognosis for Survival of Educational Evaluation at the Local School District Level?" by Carl Sewell, stresses utilizing evaluation for local educational improvement. "Is Educational Evaluation Dying?" by John W. Evans, sees the profession weathering difficult times. "Has the Profession of Educational Evaluation Changed with Changing Times?" by Daniel L. Stufflebeam, emphasizes evaluator training, the information needs of multiple audiences, and personnel evaluation. "An Analysis of Five Views of Evaluation," by Eleanor Chelimsky, comments on the preceding papers. "Educational Evaluation Will Be Reincarnated at the State and Local Levels," by Shirley A. Jackson, summarizes and analyzes audience survey responses. (BS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED251509

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy

DELIVERABLE - NOVEMBER 1984

RESEARCH INTO PRACTICE PROJECT

Evaluation Comment, 7(1), 1984

Whither Educational Evaluation: Life, Death, and Reincarnation?

Joan Herman

Project Director

TM 850 010

Grant Number

NIE-G-84-0112, P4

Center for the Study of Evaluation

Graduate School of Education

University of California, Los Angeles

The project presented or reported herein was supported pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by the National Institute of Education should be inferred.

WHITHER EDUCATIONAL EVALUATION: LIFE, DEATH, OR REINCARNATION?

Susan Klein, Eva Baker, Carl Sewell, John Evans
Daniel Stufflebeam, Eleanor Chelimsky, & Shirley Jackson

In this issue:

The authors share their thoughts about the current health and future well-being of the profession. The papers were first presented at the 1984 AERA annual meeting in a symposium organized by Susan Klein of the National Institute of Education to consider educational evaluation's health from the standpoint of five inter-related indicators -- financial support, official mandate, evaluators' status, evaluation prevalence, and evaluation utility.

Klein's introductory paper amplifies these indicators by raising a series of specific issues within each and assessing federal and national trends on them over the past several years. She concludes that there is a continuing need for federal leadership in using evaluation to improve education.

Eva Baker, UCLA Center for the Study of Evaluation, deduces from her analysis of the major indicators that evaluation stands in need of reincarnation, one which will be able to reconcile both broad policy issues and local utilization needs.

Carl Sewell, New York City Schools, reaches a conclusion similar to Eva Baker's: Educational evaluation needs to be as much concerned with local educational improvement as it is with external accountability questions.

John Evans, Educational Testing Service, though as hopeful as the other presenters that educational evaluation will continue, is less

sanguine in his assessment of its current and future health. At the very least, he suggests, the profession needs to weather some difficult times in hope of future renewal.

Daniel Stufflebeam, Western Michigan University, believes that while evaluation has changed with the changing times, greater attention needs to be placed on evaluator training, recruitment, and certification, the information needs of multiple audiences, and personnel evaluation.

Eleanor Chelimsky, General Accounting Office, summarizes each of these papers, examines their substance, and offers personal commentary. While she notes optimism in each of the papers, her own view of the general field of evaluation's future, based on her experiences in the federal government, is even brighter.

The symposium's audience was invited to assess the current status of and future trends for educational evaluation. Shirley Jackson, U.S. Department of Education, who also chaired the symposium, analyzes and summarizes audience response.

Taken together, the presentations, discussion, and audience response suggest where the educational evaluation profession has been in the past, where it seems to stand at the present, and some possible directions for the future. Such an assessment, it seems to us, is important given the need to consider the future of evaluation in light of budgetary and programmatic constraints.

EDUCATIONAL EVALUATION -- INDICATORS OF LIFE, DEATH, OR REINCARNATION

Susan S. Klein

National Institute of Education*

This paper assesses the current and future status of educational evaluation. This assessment is critical: Many believe that educational evaluation has changed substantially over the years; decreases in federal and other educational funds may have had a detrimental effect on evaluation; federal leadership in educational evaluation needs to be reconsidered in light of current and future constraints.

INDICATORS TO HELP PREDICT THE FUTURE OF EDUCATIONAL EVALUATION

The health of educational evaluation can be viewed with regard to the following five indicators:

Financial Investment in Educational Evaluation

This indicator can be gauged by the extent to which: funding or other support for conducting educational evaluations has changed; whether financial support for conducting research and development (R&D) and technical assistance on educational evaluation has changed; and the future availability of financial or other support for educational evaluation and for R&D on evaluation.

Official Status or Mandate for Evaluation

Evaluation's status might reflect the extent to which evaluation is a part of government policies, regulations, and procedures; whether these policies cover R&D on evaluation as well as the conduct of evaluation; the

*The ideas expressed in this article do not necessarily reflect the views of the National Institute of Education

existence and status of evaluation units in educational organizations; and future trends on these issues.

Status and Visibility of Evaluators

The status of evaluators might be seen in terms of the health of the profession; the extent to which educators identify themselves as "evaluators" and evaluators refer to themselves as "educational" evaluators; the availability and status of educational evaluation jobs; the health of evaluator training programs; and whether evaluators have established themselves as a profession with its own organizations, networks, norms, standards, and certification procedures.

Prevalence, Visibility, and Extent of Evaluation Activities

An indication of the health of evaluation on this measure can be seen in the extent to which people label educational improvement activities as evaluation rather than planning, research, or development; whether people see evaluation as an identifiable, priority process; the extent to which people request evaluation assistance and actually perform evaluation activities; prevalence of evaluation reports or materials on conducting evaluations; the extent to which evaluation requirements actually increase the use of evaluation at different educational levels; whether people fear and avoid educational evaluation; and whether they try to improve educational evaluation and conduct research and disseminate information on it.

Value, Utility, and Effectiveness of Evaluation in Improving Education

Evaluation's utility can be seen in terms of whether it helps improve education; its cost-effectiveness; its effectiveness relative to other educational improvement strategies; whether evaluations are used in narrow and inappropriate ways; and whether current trends are likely to influence the effectiveness of future educational evaluations.

AN ASSESSMENT OF THE HEALTH OF EDUCATIONAL EVALUATION AT THE FEDERAL/NATIONAL LEVEL

In this section I will apply the above indicators to assess federal and national trends in educational evaluation over the past 20 years with special emphasis on events of the past three years.

Financial Investment

Financial investment is a fairly quantifiable measure of educational evaluation's prosperity. But difficulties in defining and accounting for evaluation functions embedded in activities labeled in different ways make even this quantifiable indicator inexact. The dollar amounts, further, are quite small. The total funding during FY 1983 for research, development, and education (RD&E) in the Department of Education, $125 million, was less than 1% of the Department's budget, and evaluation and policy analyses, funded at $12.35 million, amounted to only 9.8% of this RD&E total. The RD&E total is expected to increase to 13% in FY 1984 (Bauer, 1984).

According to the National Institute of Education's (NIE) 1976 Databook the U.S. Office of Education (USOE) supported $12 million worth of planning and evaluation projects in 1970 and $17.4 million in 1975. However, it is likely that more evaluation activities were included in the 1976 definition of evaluation than in the definition followed by Bauer.

There has also been a shift away from large-scale program evaluations in which a small percentage of the federal program budget was reserved for evaluation. There is continued interest, however, in large-scale data collection activities such as those of the National Center for Education Statistics (NCES) and National Assessment of Educational Progress (NAEP).

Even in times of retrenchment in other federal education areas, these investments have remained fairly stable.

NIE support for R&D on educational evaluation, which exceeded $4.5 million in FY 1980, has gradually been cut in half and may decrease even more with the next solicitation of work from the laboratories and centers. Though it has been assumed that budget cuts in education and the desire to use limited resources wisely will lead to increased attention to evaluation, this assumption has not been supported by evidence relating to federal sponsorship of actual evaluations or R&D on evaluation.

Evaluation's Official Status

Federal education discretionary legislation usually carried evaluation provisions and some were quite detailed. However, the evaluation requirements for Title I and many other Department programs previously under discretionary funding were significantly reduced when these programs were incorporated in Chapters 1 and 2 of the Educational Consolidation and Improvement Act in 1981 (Burry, 1983). On the other hand, some important evaluation requirements were maintained at the local school level for Chapter 1 and in 1983 Congress asked NIE to conduct an evaluation of Chapter 1. Beginning in FY 1984, further, Chapter 2 applicants must provide for an annual evaluation of program effectiveness.

Evaluation also receives official status when program or administrative offices are designated as evaluation units. In 1969, for example, the Office of Program Planning, Budget and Evaluation was established in the USOE. The USOE began funding the UCLA Center for the Study of Evaluation in 1966 to conduct R&D on evaluation. CSE was transferred to NIE in 1972. NIE has also maintained several types of evaluation units as separate

entities or as branches within larger programs. In 1979 the General Accounting Office (GAO) assumed increased evaluation responsibilities and established the Institute for Program Evaluation. In 1981 the Department of Defense established an Evaluation Branch for its Dependent Schools.

Evaluators' Status and Visibility

In addition to funding evaluation work and requiring evaluations of federal education programs, the USOE and later NIE contributed to the professional development of evaluators through its R&D Personnel Development Program which supported evaluator training, evaluation training materials and model programs, and research and dissemination on educational evaluation.

The status of educational evaluators is also seen in the extent to which people call themselves evaluators. Of those people who responded to the 1965 National Register of Educational Researchers, 14% said that at least one of their research areas was testing, measurement, and evaluation (NIE, 1976). In 1976 9% of the AERA members listed evaluation as their primary responsibility (Egermeier, 1977). This percentage increased to 10% in 1978 and decreased to 9% in 1982 and 8% in 1983 while AERA total membership increased to 14,000 (Russell, 1983).

Other indicators of the development of the educational evaluation profession include the establishment of evaluation courses in universities, and organizations such as Division H (School Evaluation and Program Development) of AERA and the Evaluation Research Society and the Evaluation Network. The development and "permanence" of the profession may also be measured by the extent to which it has developed its own language and standards. Using this indicator, the health of educational evaluation appears robust.

Evaluation Prevalence

Evaluation activities are likely to be maintained unless they are viewed negatively or as a low priority. As described under financial investment, the federal government sees evaluation as part of its role and has continued to support and conduct program and other evaluation activities in some of its units, such as the Joint Dissemination Review Panel (JDRP) and the Secondary School Recognition Program.

Another indicator of the prevalence of evaluation activities is the amount of evaluation documents entered annually in ERIC. In 1966, 10.3% of the total ERIC entries were evaluation reports. This percentage increased to 11.7% in 1975; to 13.4% in 1980; and then decreased to 10.7% in 1983. Similar trends appear when only USOE/Department or NIE evaluation entries are examined.

From FY 1980 to 1983, NIE project officers indicated that 10.5% of their projects' activities were for evaluation compared to an average of 8% for policy analysis and 20% for applied research. Over the years a small subset of these projects were for research on evaluation.

Since the early 1970s various independent organizations have also conducted evaluations for the federal government and other educational institutions. Nationally, by 1978 16% of the R&D organizations who responded to an NIE sponsored survey listed evaluation as one of their principal activities (Campbell & Brown 1982). It is possible that this percentage is much lower now.

Evaluation and Educational Improvement

This indicator is difficult to use, and complicated by the shifting popularity of certain kinds of evaluation over the years. In the early 1970s, for example, much federal attention was given to using evaluation

for educational accountability and establishing behavioral objectives. Later the emphasis was on using evaluation for program decision making and Brickell's Data for Decisions (1974) served as a guide for federal evaluation reports. Today there is as much emphasis on using evaluation information to increase people's understanding as there is on their making decisions on the basis of an evaluation report's recommendations.

The Department and national policy groups such as the National Academy of Sciences have assessed the value of some evaluations and have supported evaluations of evaluations. There is also increased attention to the cost effectiveness of evaluations. NIE has taken its initial mandate to improve the educational R&D system seriously, and supported a wide variety of projects to develop and refine evaluation tools and practices.

Similarly, NIE has supported research studies on the role of evaluation in educational improvement ranging from an analysis of its use in exemplary programs which have received JDRP approval (Klein, 1984) to how school administrators acquire and use knowledge from evaluations. Although many evaluations have not been as useful as expected, there is evidence to show that their utility is increasing (Stalford, 1983) and that evaluations can have direct value in improving education, although there has been relatively little emphasis on or success with comparative evaluations. There has also been little information on the differential effects of the treatments or programs on girls and boys and majority and minority students, despite the Department's requirement that all grants do this and the concurrent emphasis of many federal programs on providing equal educational opportunities (Stalford, Millsap, et al, 1981).

A PROGNOSIS FOR EDUCATIONAL EVALUATION AT THE FEDERAL LEVEL

Based on historical federal involvement with educational evaluation, some evidence that evaluation is one effective educational improvement strategy, and evaluation's establishment as a profession, I predict that evaluation will continue to live.

Certain evaluation activities and needs are likely to remain at the federal level; other evaluation functions are likely to be reincarnated in altered forms. Many evaluations will continue to be "mainstreamed" and become invisible components of other educational improvement strategies. Other evaluations may focus on products and practices rather than on complex federally sponsored programs. It is also likely that evaluation activities will become more cost effective through greater use of computers. However, in times of tight money, it is likely that discrete evaluations will be less frequent, smaller, and more focused.

Defining and assessing federal involvement in educational evaluation is already difficult and will be even more challenging as evaluation continues to change. There will be a continued need for federal leadership to ensure that these changes are beneficial and help improve education.

REFERENCES

Bauer, L. Department of Education fiscal years 1983 to 1984 research, development, demonstration, evaluation, and policy study activities. Washington, D.C.: U.S. Department of Education, 1984.

Brickell, H.M., Aslanian, C.B., & Spak, L.J. Data for decisions: An analysis of evaluation data needed by decision makers in educational programs. A report to the National Institute of Education. New York: Policy Studies in Education, 1974.

Burry, J. Federal regulations affecting local evaluation use. CSE Report No. 222. Los Angeles: UCLA Center for the Study of Evaluation, 1983.

Campbell, P.B., & Brown, M.V. Survey on the status of educational researchers. A report to the National Institute of Education. Washington, D.C.: 1982.

Egermeier, J.C. Participation of women and minority subgroups in research-related activities in the field of education. Washington, D.C.: National Institute of Education, 1977.

Klein, S.S. Is R&D a key component of effective educational programs? Washington, D.C.: National Institute of Education, 1984.

National Institute of Education. Databook: The status of education research and development in the U.S. Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1976.

Russell, W. Demographic characteristics of the AERA membership. Educational Researcher, August-September, 1983.

Stalford, C. Annual report, evaluation team: Testing, assessment, and evaluation. Washington, D.C.: National Institute of Education, 1983.

Stalford, C., Millsap, M.A., Gold, N., & Antonoplos, D. Research area plan in program development and evaluation. Washington, D.C.: National Institute of Education, 1981 .

IS THERE A PERCEIVED OR REAL METAMORPHOSIS OF PURPOSE FOR EDUCATIONAL EVALUATION?

Eva L. Baker

Center for the Study of Evaluation

I have been blessed with a range of evaluation experience including studies of national policy import, of state educational reform, of tricky and emotional local programs such as involuntary and voluntary desegregation, and of special programs in post-secondary, public school, and the private sector.

This experience, which has been both exhilarating and painful, makes me reluctant to predict educational evaluation's health as if it were one entity with a well-bounded field. Rather, I see evaluation as taking many guises. I predict that some will continue to live, I support euthanasia for others, and I fervently hope that some will transmute and come back to us in a better form.

With these remarks as background, let me turn to the current status and future possibilities for evaluation.

Financial Investment in Evaluation

Financial investment in evaluation has changed, but not as drastically as we expected. The source of the change is unclear: Have reductions in evaluation been caused by financial cut-backs of the last three years, or do they stem from dissatisfaction with evaluation itself? In fact, agencies with money seem to be doing more evaluation now than ever before, which suggests that any market decline we are experiencing is not entirely a matter of disdain for the product. Rather, program officers facing a

financial crunch seem to follow a simple plan: Spend money on the program itself rather than on studying the effects of the program.

Evaluation's emphasis, too, has changed, with policy makers focusing more now on issues of system performance and achievement. At CSE, we see this emphasis in requests to design management information and pupil performance systems. These systems are intended to provide continuing sources of information about a range of programs, which is quite different from earlier emphasis on one discrete evaluation of a particular project.

Official Status or Mandate for Evaluation

With regard to status, evaluation continues to be institutionalized. Offices of evaluation still exist, and substantial investment is provided for evaluation activities. Most of these activities, however, emphasize hard indicators of effectiveness to the detriment of those softer indicators which are data-rich, have an intensive base, and are locally relevant. For example, in our evaluation work with a large school district over the last three years, pressures have been mounting to develop better indicators of pupil performance. Yet the findings most interesting to clients and those resulting in program change have been those stemming from intensive studies of a few schools or classrooms.

The status of evaluation units is hard to gauge. A unit like the GAO, which has a sharply-defined federal role and whose findings are connected to real and public consequences, has high status. In many situations, however, I believe the status of evaluation units depends upon how well they serve the policy and political needs of the superordinate institutions, and this status varies with the institution and the competency of the evaluation unit itself.

Another indicator of the health of evaluation is the attention lavished on studies putatively assessing the status of American education. If press attention is a measure, and if these studies can be included in a broad definition of evaluation, then the field still has some kick left in it.

Status and Visibility of Evaluators

The status of evaluation is directly related to the status of evaluators themselves. Evaluators seem to have lost some of their celebrity status, judging from attendance at AERA sessions. In the 1984 program, as was the case last year, symposia suggest that the patina is wearing and that we are concerned with whether anything we do in research and evaluation helps improve the state of education. Almost all of education has been shaken by the mediocrity challenge, and the status of evaluators, as part of the larger profession, is under scrutiny.

In regard to other indicators of the health of the profession, Susan Klein suggests we address evaluators as a generic class versus an educational specialty. People in schools of education focus principally, and logically, on educational phenomena, and have special expertise for understanding and clarifying educational programs and systems. Those operating out of social science disciplines see their arena more broadly. Even so, at UCLA we have conducted evaluation activities dealing with technology, with management, with organizations, and with the delivery of services in health, housing, and social services.

Taking up another of Susan Klein's prompts, the job market, as I gauge it from placement of our students, is fine. But as students enter the job market, they are likely to encounter roles and positions which include

evaluation responsibilities rather than slots exclusively labeled as "evaluator".

As another indicator of health, our evaluation training program at UCLA is doing well. We have better students than ever, and those choosing evaluation, when its status has been the topic of increased public scrutiny, seem to be deeply committed to improving the state of education.

A final indicator of health has to do with the networks, organizations, and standards spawned by the profession. Work on standards, led admirably by Dan Stufflebeam, has been completed. Organizations and networks are emerging faster than we can give them names.

Prevalence, Visibility, and Extent of Evaluation Activities

How visible ought evaluation to be? To the extent that one believes, as I do, that the usefulness of evaluation rests on its regular integration in the planning and conduct of educational programs, then its identity should be subordinate to the programs.

Such a view is not widely prevalent. While evaluation activity has subsided modestly, the character of evaluation has become more local, more interactive, more political, and more targeted. Local control of educational services has resulted in evaluations that have shorter time-lines for delivery, whose scopes exceed what financial resources permit, and which leave limited opportunity for innovation. Requests for proposals for evaluation seem to follow the line of much contract research. Government personnel are specifying not only the information they want but also the methods and tactics they view as acceptable. Deliverables are more tightly phased, and the result overall is not good for the field. Because of market pressures, however, many evaluation specialists still propose and conduct such studies. I am concerned that such processes, should they

continue, will drive the serious, reflective professionals out of evaluation leaving only those willing to comply with overspecification.

Sometimes onerous for the evaluator, I think evaluation is viewed as generally unpleasant and unproductive by most who undergo it. Such reactions can be avoided by deliberate, careful involvement of all parties, as described by Patton (1982), Bank and Williams (1981), and Baker (1984). But such activities take time and other resources and run flat into the trend for shorter time lines and minimal funding.

Research on evaluation goes forward, but in less volume. However, research on methods to improve evaluation's real-time practical import, and descriptive studies of evaluations as they occur, as well as more theoretical advances, provide great benefit for the field. If one's goal is, as mine, to produce more effective programs using cost-sensitive procedures, then I believe research on evaluation needs to be continued.

Value, Utility, and Effectiveness of Evaluation in Improving Education

The emergence of evaluation utilization as a field of study raised questions about its value. While on the one hand the research evidence for evaluation utilization is less than heartwarming, on the other hand, we are experiencing at UCLA a happy trend of having real policy decisions, at school districts, and to some degree at the state level, depend in a surprisingly linear way on our findings.

Alkin and Solmon (1983) and Catterall (1983) at UCLA, Levin (1983) at Stanford, and others have been studying costs of evaluation. Also at CSE, Herman and Dorr-Bremme (1983), Williams and Bank (1981), and Sirotnik, Burstein, and Thomas (1983) have been addressing the utilization of testing at national, district, and school building levels. A next step is to

examine the cost effectiveness of various strategies, although determining the cost effectiveness of evaluation depends upon a number of variable factors; e.g., the character of the program, the immediacy of the need, and the options one has available to improve the service.

My bias is for formative evaluation that leads to the improvement of program effectiveness as measured by indicators of performance, satisfaction, and cost, as well as of implementation and service delivery. My concern, however, is that the methods we have so carefully adopted, and adapted, from social science research often corrode and inhibit the development of productive programs. When people worry about who's watching them too much, using methods and indicators marginally relevant, they tend to take conservative, status-sustaining postures. We need less of that.

We are developing a study group at CSE to bring together outstanding methodologists and evaluation thinkers to ask questions about these issues. Perhaps our psychometric dependence can become less dominant so that our decisions in evaluation tactics follow the problem rather than the methodology. Perhaps we can find new ways to describe the quality of school performance beyond annual rankings on some standardized test. I believe there is a fair chance of this happening because evaluation seems to be sufficiently mature to leave behind the "opposing camps" -- step-wise regression versus ethnography -- atmosphere of the past, to take a problem focus, and to invent or recombine strategies and tactics to produce a more effective enterprise.

A Model of Top-down, Bottom-up Evaluation

Though I have often chanted the litany of evaluation models, I think there is room for a new one, one which maximizes the utility of evaluation. The model recognizes that the needs for evaluation, from a policy

perspective, push us in one direction, but our knowledge about evaluation utilization pushes us in quite an opposite way. Policy evaluation comes from the top down, utilization springs from the bottom up, and so I propose a "top-down, bottom-up" model in which both purposes and uses are combined.

Top-down evaluation is exogenous to the data providers. Data are summarized for broad decisions and to meet public needs for accountability. Thus, it is important to be able, in a school district, for example, to compare performance across children, schools, and maybe even teachers, to decide on what is working. Such an approach requires that we use general measures, comparable sampling, and other methods reflecting the strength of social science.

But no matter what decision is forthcoming, it is the people at the schools, in classrooms, who have to be convinced, persuaded, and encouraged to change. Research strongly suggests that the school is the appropriate unit of change, even though policy making occurs outside of the school.

How can we maximize utility at the school level, from the bottom up? We can go to a growing knowledge base on school and teacher needs in this area. A few examples: Teachers need timely information so that they can provide help to students; they also need information which is relevant and sensitive to their instructional programs and emphases. Schools are different, and common measures are often inappropriate across sites. Data-users need to feel safe about providing information, and need to feel that descriptions of what they do are fair. There need to be incentives to change, and evaluation strategies need to consider stakeholders and ownership.

I would hope that evaluation would be reincarnated into a more useful activity so that:

- The locus of evaluation is placed at the unit of change, at the school.
- Evaluation systems are built to secure a place for both top-down accountability and policy concerns and bottom-up utilization issues, preserving sufficient local options to address issues of particular interest to individual schools.
- Quality of data issues are attuned to the facts of utilization; e.g., we know teachers use many sources of information, and so perhaps the reliability of any one source can be safely reduced.
- The system avoids data innundation.
- The system provides for phasing in and out of information.

The research questions related to such a model abound. They include discrepancy or conflict resolution among data sources, mixes of methods and measurement, analysis and interpretation issues. There are also questions reflecting technology, user-friendliness, and the realm of incentives, commitment, collaboration, disclosure, and change.

The end might be not a single operating system but a set of principles and methods that enable us to change evaluation from something that a few do to many to a process that:

- answers questions for a variety of audiences
- includes a range of measures
- has real-time utility
- conserves time of teachers and students
- can be demonstrated to improve practice.

REFERENCES

Alkin, M.C., & Solmon, L.C. (Eds,). The costs of evaluation. Beverly Hills, CA: Sage Publications, 1983.

Baker, E.L. Tests and the real world: The problem of utilization. In S.J. Hueftle (Ed.), The utilization of evaluation. Minneapolis, MN: Minnesota Research and Evaluation Center, 1984.

Bank, A., & Williams, R.C. (Eds,). Evaluation in school districts: Organizational constraints. Los Angeles: CSE Monograph No. 10, UCLA Center for the Study of Evaluation, 1981.

Catterall, J. The cost of instructional information systems: Results from two systems. Los Angeles: CSE Report No. 208. UCLA Center for the Study of Evaluation, 1983.

Herman, J., & Dorr-Bremme, D. Uses of testing in the schools: A national profile. In W.E. Hathaway (Ed.), Educational testing in schools: New directions for testing and measurement No. 19. San Francisco: Jossey-Bass, 1983.

Levin, H.M. Cost-effectiveness: A primer. Beverly Hills, CA: Sage Publications, 1983.

Patton, M.Q. Practical evaluation. Beverly Hills, CA: Sage Publications, 1982.

Sirotnik, K., Burstein, L., & Thomas, C. Systemic evaluation. Los Angeles: UCLA Center for the Study of Evaluation, 1983.

Williams, R.C., & Bank, A. Uses of data to improve instruction in local school districts: Problems and possibilities. In C.B. Aslanian (Ed.), Improving educational evaluation methods. Beverly Hills, CA: Sage Publications, 1981.

WHAT IS THE PROGNOSIS FOR SURVIVAL OF EDUCATIONAL EVALUATION AT THE LOCAL SCHOOL DISTRICT LEVEL?

Carl Sewell
Brooklyn (NY) Public Schools

In this presentation I want to share with you an assessment of evaluation from the standpoint of a local district practitioner. My assessment will generally follow the indicators that Susan Klein previously described.

Official Status and Institutionalization

One view of the status of evaluation derives from examination of how educational policy makers -- such as school board members -- behave with respect to evaluation. From my own experiences with such policy makers, I conclude that they tend to be suspicious of educational program evaluation with few actually understanding its processes.

In the local school district context, policy makers seem primarily to place serious consideration or priority on evaluation when there is a sense that something is wrong in the school system and, from this perspective, to view evaluation as a potentially punitive process.

Evaluation also gets attention when it is necessary for satisfying requirements for continued funding. However, such required evaluation is rarely seen as a tool for local decision making in areas such as program improvement, either from the perspective of the local school district or the funding source.

The status of evaluation at the local district level, then, often grows out of an arena reflecting both punitive and political concerns. To the extent that accountability concerns arise within this context, then (1) the over-riding concern for the local policy maker is to show how much his or her schools have "improved" since he or she has been in charge of

the educational system; (2) summative, end-of-program evaluation data, most frequently in the form of norm-referenced achievement test data, are emphasized; and (3) there is little, if any concern, for formative evaluations that provide crucial feedback for educational planning, monitoring, and improvement of instructional programs and services.

Given the local policy maker's perception of evaluation -- a potentially punitive and highly political endeavor -- it is not surprising that local institutionalization of evaluation is limited, even in the larger school districts with a history of maintaining a "research" or "evaluation" unit in the organization. This limited institutionalization of the evaluation process is also seen in how resources are expended for evaluation. Resources given over to evaluation are usually sufficient only to satisfy legal requirements, to serve purely political needs, or to counteract public criticism. It is rare for evaluation to be designed and used for ongoing program development and improvement. In fact, I believe most local school district policymakers, given their limited understanding of the evaluation process, would view its proactive use as a fiscal luxury.

Status, Ability, and Training of the Evaluator

The status of evaluation personnel in the local school district organization varies from setting to setting. For example, in those districts where there is a distinct evaluation unit in the organization, the head of the unit normally has administrative and supervisory status. While this status usually requires appropriate educational credentialing, it generally does not demand extensive training or experience in research and evaluation, and so the unit head often serves as broker and coordinator of

specialized research and evaluation activities, sometimes carried out by unit personnel, sometimes by consultants.

Now, successful brokering and coordination of services can be an important administrative function. For example, a large and complex evaluation may be facilitated when an evaluation unit head knows how to assign, coordinate, and judge the worth of a variety of evaluation activities. However, if brokering reflects inability to perform the evaluation services rather than the need to coordinate many services, it diminishes the unit's status and credibility. To increase their status and credibility, it is essential that the evaluation unit head and staff have a strong background in research and evaluation, even if they rely heavily on consultants. Without this expertise, they will be hard-pressed to mount technically and contextually sound evaluations and to translate their results into effective strategies for improvement. Unless evaluation unit heads and their staff come to acquire this expertise, it is likely that they will continue to enjoy limited organizational status. Their status may increase, however, if they can design useful evaluations which provide information of value in educational planning and problem solving.

Financial Support

In light of my remarks about evaluation's institutionalization and status, and especially because its local application is dominated by reactive considerations, it is not too surprising that evaluation, in the local school district, does not receive high fiscal priority. When budgets are seriously constrained, the major portion of program evaluation is contracted to outside consultants who are restricted to required evaluation tasks. In order to receive the budgetary consideration that it deserves,

evaluation must come to be viewed as a means of generating information which can do much more than meeting funding or political matters.

Therefore, just as the providers of evaluation information (the evaluation unit staff) need to acquire greater evaluation expertise, so must the users of evaluation information (local policy makers) come to acquire greater understanding of the evaluation process and its potential contribution. Without such understanding, it is doubtful that policy makers will come to see evaluation as fundamental to the conception, planning, and delivery of effective and fiscally sound educational programs, and as a means of communicating understandable information to their communities.

The problem I am outlining here, further, is exacerbated by the role played by school district administrators as they interact between evaluators and local policy makers such as the school board. Let's assume that we can realize greater expertise within our evaluation units. Let's also assume that we can increase policy makers' understanding of and receptivity toward evaluation information. These gains will fall by the wayside unless school and district administrators come to understand the administrative uses of evaluation in planning, developing, and implementing programs and reporting their results to the board. If evaluation is not seen by administrators as an important administrative tool, then it is unlikely that it will come to be viewed as a high priority by the board. I see a great need, then, to increase the skills of local administrators in the utilization of evaluation information to enhance the quality of their leadership and decision-making. Without such utilization, evaluation will continue to recieve limited support.

Prevalence, Value, and Extent of Evaluation Activities

Given increasing demands for accountability and continuing fiscal constraints upon public funding for education, it seems reasonable to predict that evaluation, at least in the restricted sense I have described here, will carry on. But as my previous remarks should have already suggested, I do not believe that such continuity is sufficient; to increase its value, evaluation must outgrow the "required" stage and come to realize its potential as a critical ingredient in the improvement of our educational systems. I have tried to convey a sense of some of the building blocks that will be required for evaluation to realize this potential.

In its present manifestation, I believe that evaluation already has some proven value for the decision makers (policy maker or adminstrator) at the local school district level. It needs to be broadened enough in scope to tap its full potential as a decision making tool.

An Outlook

What's my prognosis for the future of educational evaluation? With shrinking budgets, increasing demands for accountability, and requirements for evaluation tied to program funding, evaluation -- at least in a limited form -- will continue to be a prevalent practice at the local school district level. The issue, it seems to me, is to move beyond that limited form. That is, we as educators need to awaken this "sleeping giant" and harness its potential as a decision-making tool that helps us plan and implement organizational strategies and instructional methods to improve our educational programs.

IS EDUCATIONAL EVALUATION DYING?

John W. Evans

Educational Testing Service

In this presentation I have been asked to address the question: "Is Educational Evaluation Really Dying?" The answer is, if it isn't dying, its precarious condition is grounds for the most serious concern.

But to understand where we are and where we might be headed, we need to look first at where we have been -- not least because compared to looking ahead, looking back is very exhilarating.

A Brief Look at the Past

The brief but scintillating history of educational evaluation constitutes a truly remarkable chapter in the history of social science. In the decade spanning the late sixties to the late seventies, the expansion of educational evaluation is little short of astonishing. The funds spent on educational evaluation went from the hundreds of thousands to the hundreds of millions. The number of people who could lay reasonable claim to the title of educational evaluator went from a small handful to several thousand. The volume of studies, evaluation units in government agencies, private research firms, academic programs in evaluation, professional societies and journals devoted to evaluation, legislatively mandated requirements for evaluation, and use of evaluation findings by legislators, managers, and educators increased dramatically during this brief period.

There were many conditions underlying this explosive growth, but two in particular were overridingly important. First was the political tide of social reform seen in federally supported programs attacking a broad array

of inequities in education, health, housing, jobs, civil rights, and the environment. Second was the determination of a small pioneering band of federal planners, budgeteers, and social scientists to launch reform efforts which were based on research, and which were constantly assessed and improved by means of rigorous evaluation in order not merely to avoid waste, but to avoid dissappointment and ensure success.

Many of the large federal evaluations carried out during this era were accompanied by intense controversy over the distastefulness of their findings and the validity of their methods. Out of all this activity and the accompanying debates, the contributions that evaluation and evaluators made to this entire period were remarkable -- in the areas of evaluation design, measurement, and statistical analysis, in the conduct of large and complex national evaluations, in compromising the conflicting interests of the various stakeholders in evaluation studies, in bringing federal, state, and local levels of government together as partners in the evaluation process, and in goading political and institutional decision makers to take account of the findings of evaluation studies.

The Current State of Affairs

Despite the flourishing of evaluation during this brief period and the remarkable achievements it has registered, it is clear that program evaluation in general, and its educational sector in particular, are now in a deep depression, and their future is seriously problematic. The indications of decline are manifest in the closing down of private research firms, the dismantling of government evaluation units, the phasing out of academic programs, and the virtual disappearance of job openings. What was

once a bountiful flow of funds for the support of evaluation studies and staffs has dried up to a mere trickle.

Why has this happened? It has not happened because of any of the numerous internal shortcomings and flaws -- lateness of reports, poor designs, or jargonish prose -- which evaluators themselves have been so candidly self-critical about. Nor is it due to the fact that Congress and other levels of government were too politically motivated to utilize evaluation findings. (See Leviton & Boruch, 1983, for documentation of the extensive use of educational evaluation.)

The sudden, and unpredicted decline in evaluation activity is the consequence of much more fundamental conditions, namely, the sharp decline in federal financial support for evaluation studies which, in turn, is the result of the cutback and consolidation in federal social reform programs. For it was, after all, the federal programs and the resources to evaluate them which were the main engines of bringing the extensive evaluation enterprise into being. Unfortunately, that enterprise appears not to be self-sustaining once those large federal programs and resources have been reduced.

An Outlook for the Future

What is the outlook for the future? How long will it be before we can expect the halcyon days to return?

On the surface it seems logical that evaluation would be a natural priority of a conservative administration interested in reducing waste and streamlining government. Indeed, as I have observed elsewhere, it is an article of faith in the policy research field that the most opportune time for policy analysis and evaluation is during periods of retrenchment,

because "reductions will force hard analysis and rational trade-offs that are not required during periods of expansion. My own experience is that this principle is totally wrong. It is during periods of expansion... that the choices of what to do and how to do it are much more open to the influence of research and analysis." Under retrenchment, programs are cut only as a last resort, "and where the cuts are made is based almost entirely on political strength or weakness, not on considerations of relative priority or effectiveness" (Evans, 1979).

More important, in assessing evaluation's future we should not deceive ourselves that current cutbacks in federal social reform programs and their evaluations are short-term effects. The changes which underlie the current situation are far more fundamental. Indeed, we are witnessing a major political and economic watershed. Mr. Reagan is the first president since Roosevelt began them to propose cutbacks in social reform programs, and to carry them out with broad public support. One of the long standing cornerstones of federal policy -- assuming responsibility for social reform and supporting programs that deal with social problems -- is eroding away. The country has become disillusioned because the problems did not yield to easy solutions, and the fiscal burden of continuing the efforts has become unacceptable as deficits soar and their economic consequences become more threatening. The federal commitment has fallen back from expansive, idealistic innovation to irreducible, obligatory maintenance.

The social, economic, and political changes which have created this new scenario are profound, and are not likely to be reversed in the next presidential election, or indeed soon thereafter.

The major federal social reform efforts, in the form of large categorical programs, were the main stimulus to the development and expansion of

the program evaluation industry. Their demise carries evaluation with them, and the poor prospect for the re-emergence of such efforts also spells a dim future for evaluation as we have known it.

This is a painful assessment to make, especially for those of us who were the midwives of the evaluation enterprise, and who nourished it through its flourishing adolescence. I do not want to be the one to read its obituary, and so I will conclude by straining to find a few silver threads in the dark clouds.

Evaluation's Contributions

The most positive thing that can be said is that the legacy of this productive if brief period of evaluation is manifold and systemic. Its effects will have far reaching and long lasting influences throughout education and social science.

First, there is, of course, the stimulus that this intense period of evaluation activity has provided to the advancement and improvement of evaluation methodology -- in measurement and analysis, but in particular to the development and application of quasi-experimental evaluation designs (see, for example, Campbell, 1969).

More important has been the empirically based additions to the education knowledge base. Through evaluation studies we have produced important new knowledge relating to the mechanics of the educational process -- knowledge, for example, on the effects of class size, the verbal abilities of teachers, the important variables in effective compensatory education programs, and so on.

At a much higher level, evaluation studies have added to the repository of basic propositions about the effectiveness of educational

strategies. Many of these propositions have by now become such common intellectual coin that we forget how recently it was that we proceeded on different assumptions:

- The cumulative results of literally hundreds of evaluation studies, beginning of course with the Coleman Report, have finally laid to rest the long standing assumption among educators, legislators, and the public that resource inputs and increases in them will automatically give rise to educational outputs.

- A major if chastening conclusion which must be read from the sweep of evaluation literature is that the ability to get a program effect is difficult to the point of being unlikely. The literature, not only from education but from the other social reform fields of employment training, criminal justice, and mental health, indicates that when rigorous evaluation methods are applied, most programs and most treatments are found to be largely ineffective.

- But that same literature also indicates that, contrary to the doomsayers who have concluded that large educational programs can never succeed, highly effective, school-based programs can be devised that can be shown under conditions of rigorous quasi-experimental evaluation to produce substantial effects on both achievement and motivation.

- The accumulated evaluation literature has documented the overwhelmingly dominant influence of out-of-school factors, and thus hopefully helped to move educators, policy makers, and legislators away from the fruitless course of attempting to accomplish miracles with the limited influence that the school and its factors provide.

But the most important contribution of educational evaluation has been its institutionalizing an evidentiary way of thinking about educational programs and policies, previously -- it must be said -- a somewhat foreign cognitive mode for many educators. Now, among educators at all levels, one increasingly encounters a healthy skepticism about newly proffered education programs and solutions, and a demand for evidence of effectiveness.

In this same vein, the public's increased awareness of the outcomes and methods of educational evaluations explains demands for hard evidence on achievement, academic standards, and teacher performance.

An Agenda

These achievements are impressive and suggest that the legacy of evaluation will indeed carry on in reincarnated form. As to how that concretely might be done, Howard Freeman (1983) has recently put forth an important agenda relevant to the new evaluation scenario and current fiscal and political circumstances. He calls for:

- Studies which estimate the net impact of cutbacks in social programs in terms of real cost savings and consequences for the target populations and the society at large.
- Implementation of social experiments to provide services at reduced cost through alternative and innovative modes of program delivery.
- Examining the appropriateness of the objectives of current social programs and their relative priority.
- Assessing the effectiveness and efficiency of established programs in comparison with alternative interventions.
- Developing procedures for ensuring accountability and maximizing the impact of programs placed under block grants.

Carrying out the evaluations embedded in this agenda would certainly provide useful -- indeed invaluable -- information in the current fiscal and political setting. However, as Freeman himself has noted, "it will be necessary to convince the current administration and conservative legislators that program evaluation can contribute to optimizing the return from public expenditures for social programs..."(Freeman, 1983).

Where all this leaves us, I'm afraid, is that while educational evaluation may not yet be ready for the undertaker, it would have difficulty buying life insurance. Nevertheless, there is hope in the fact that the utility of evaluation cannot be permanently overlooked. The record of accomplishments is impressive, the professional commitment and vitality of evaluators remain strong, and the need for rigorous evaluations, from the

federal government down to the smallest school building, remains high. Therefore, despite the current recession of opportunities and support, evaluators, and those who understand the ways in which educational evaluation has already contributed to making educational policies and programs better, have no choice. They must weather these difficult times and rededicate themselves to the sustenance of this important enterprise.

REFERENCES

Campbell, D. Reforms as experiments. American Psychologist, 1969, 24, 409-429.

Evans, J.W. Thoughts on the status of applied social research. In C.C. Abt (Ed.), Problems in American social policy. Cambridge, MA.: 1979, p. 170.

Freeman, H. A federal evaluation agenda for the 1980's: Some speculations and suggestions. Educational Evaluation and Policy Analysis, 1983, 5(2), 186.

Leviton, L.C., & Boruch, R.F. Contributions of evaluation to education programs and policy. Evaluation Review, 1983, 7(5), 563-598.

HAS THE PROFESSION OF EDUCATIONAL EVALUATION CHANGED WITH CHANGING TIMES?

Daniel L. Stufflebeam

Western Michigan University

In this presentation I examine whether educational evaluation has changed with the changing times. I'll begin with a framework for considering changes in the professional character of evaluation. Then I'll comment about what I see to be the main trends. Finally I'll offer some thoughts about what lies ahead and a partial agenda for strengthening the profession.

A FRAMEWORK FOR CONSIDERING THE EVALUATION PROFESSION

In developing professionalization variables for evaluation I drew on Boulding's ( ) analysis of the critical characteristics of a secure or mature profession -- an adequate historical record, a knowledge base covering the total field, clearly defined and relatively simple structures and relationships within the field, and a concern with widespread and recurrent events. After applying this analysis to evaluation, the resultant framework suggested that educational evaluation may be studied in terms of its substance, clients, practitioners, and formal structures.

WHAT APPEAR TO BE THE MAIN TRENDS.

Over the past twenty years there has been great expansion in the four indicators of the evaluation profession.

Substance

From the middle 1930s until the middle 1960s a narrow view of educational evaluation prevailed. That view saw evaluation as determining

whether specified objectives had been achieved, and the preferred methods included behavioral objectives, standardized tests, and experimental designs.

The 1960s saw alternative conceptualizations. In his landmark (1963) article, Cronbach charged that comparative studies and norm-referenced test results were not very informative or useful. He suggested that formative studies and specific item analyses would be of much greater service in efforts to improve education. About 1966, evaluators who had been trying to implement the classical evaluation views in federal projects found that Cronbach's points were valid; criticisms and proposals for reform grew, and alternative conceptualizations and methods appeared in the literature (Scriven _____, _____; Stake _____; Guba _____; Owen _____; Wolf _____; Stufflebeam _____; Smith _____).

Though the early 1970s were characterized by controversy and dissension concerning these conceptualizations, the differences in proposals were more apparent than real. The Joint Committee on Standards for Educational Evaluation (1981) provided concrete evidence of underlying harmony across different conceptualizations of evaluation, and reached agreement on a basic set of standards for judging evaluation work. A second committee, appointed by the Evaluation Research Society ( ), developed a set of standards substantially in accord with those of the Joint Committee.

Clients

In the early 1960s the main client group included the sponsors of the large national curriculum development projects. It was generally assumed that their questions could be summed up in the rationale underlying a good

experimental design; i.e., to what extent is one program superior to another one in producing a desired outcome?

This situation changed dramatically in 1965 when the Congress mandated evaluations of projects funded under the Elementary and Secondary Education Act (ESEA). Seemingly, the audience for evaluation became much broader and the questions more complex. However, initially the main audience turned out to be federal bureaucrats trying to interpret and enforce the Congressional mandate. They were very confused about what questions should be addressed and they left it to school districts to set and respond to the questions they thought would be of interest. But when the U.S. Office of Education evaluators tried to aggregate the local school district reports, the results were an embarrassingly bad report that didn't address any group's questions. All who were involved learned the necessity of differentiating among audiences and of identifying and directly addressing their different questions; current interactive approaches to evaluation address this problem.

Educational evaluators also increased their efforts at the state and local levels during the late 1960s. State education departments had strengthened their evaluation capabilities through support from Title V of ESEA; the superintendents of many large and middle sized school districts began to recognize the necessity of sound evaluations of their programs; federal grants had enabled these districts to greatly strengthen their evaluation capabilities.

As states took on increased responsibilities for evaluation, they too found that identifying and addressing the questions of their audiences were very important but difficult tasks. New state-level teams tried to

stimulate interest in questions to guide evaluation and research work; in a sense they were their own clients.

The large urban school districts are responsible for much of the progress made in increasing evaluation's utility. They had convened sizeable staffs of highly trained evaluators and, because of their many external and internal pressures, they had a continuing need for information of use in guiding decision making and meeting accountability requirements. However, these districts have found that a team of specialists can only serve some of the audiences and some of their information requirements. With few exceptions, the tradeoff has been to serve the granting agencies and the superintendent first, and only infrequently to provide reports aimed at teachers and principals. We still need improvement in how evaluation identifies and serves the multiple audiences in school districts.

Audiences outside education are developing an interest in evaluation, and some educational evaluators have found that their services are being sought by groups in other fields. This raises a question as to the pros and cons of attempts to maintain a profession of educational evaluation or to broaden the domain to encompass an expanded array of social and educational services.

Practitioners

Until the late 1960s educational evaluations were mainly conducted by a few measurement and statistics specialists. Then, in the late 1960s, when school districts throughout the U.S. found they had to evaluate their Title I programs, the ranks of educational evaluators were greatly expanded by drawing in educational researchers, counselors, and psychologists -- anyone claiming some expertise and/or willingness to take on the

assignment. The result was a great deal of poor work (Guba,), reflecting a lack of pertinent evaluation training and a lack of valid evaluation theory and methods in which to be trained.

The 1970s marked an increase in people trained specifically to do evaluation work. Some came from university training programs; others obtained specialized training in workshops such as those sponsored by AERA over the past 15 years; researchers from outside education joined the ranks of educational evaluators. The educational evaluators of the 1970s represented a significant advance in education's capacity to evaluate its programs.

In the 1980s education has lost many of its most able evaluators. In some cases the evaluators have been moved into different positions; in other cases, they have moved outside education to take evaluation positions elsewhere. This mobility might not be bad and it might even be desirable were it not for the widespread failure to replace departing evaluators.

Formal Structures

Prior to the middle 1960s there were essentially no formal structures to foster the professionalization of educational evaluation. There was no professional society; the literature was mainly restricted to a few booklets by Ralph Tyler (__ , __); the main training program seems to have been the one Tyler directed at the University of Chicago in the 1940s and 1950s; there were no standards for judging evaluation work, except for the Standards for Educational and Psychological Tests ( ); there weren't any programs for certifying or licensing educational evaluators. Now, there is evidence of growth in relation to each of these aspects.

In the middle 1960s, for example, the American Educational Research Association became involved in examining and fostering improvement in evaluation practice. The Association sponsored many symposia and workshops on evaluation. It created a division for school evaluators. It developed a journal of Educational Evaluation and Policy Analysis.

In addition, several new evaluation societies were established, including the Evaluation Network, the Evaluation Research Society, the May 12th Group, and the Joint Committee on Standards for Educational Evaluation. The membership of these groups numbers in the thousands, but is down from its peak in the early 1970s of more than 5,000 members.

Anyone who has been trying to keep up with the literature of educational evaluation over the past 15 years is well aware that there has been an explosion in this area. We now have a wealth of published materials, and the problem, say from the standpoint of a university professor, is not to find relevant course material but rather to keep up with what is available.

There has also been some improvement in the training of evaluators, but there has also been a loss in relation to the gains made in the late 1960s and early 1970s. Then a number of universities, AERA, the U.S. Office of Education, and some state education departments and regional educational laboratories provided both inservice training and degree programs, and these programs turned out many of the persons who later did outstanding evaluation work. But in recent years I believe there has been a general decrease in efforts to recruit, train, and place graduates into educational evaluation work. This decrease parallels the downturn in the economy, the decrease in funds for educational evaluation work, and the

federal government's severe cutback of support for evaluation training. Also, the track record, overall, in recruiting and training minority students for work in evaluation has not been good.

Twenty years ago there were standards for evaluating educational and psychological tests but not for judging and guiding evaluations, per se. But the field has advanced a long way in this regard, as seen in the Joint Committee standards and those of the Evaluation Research Society previously mentioned.

My final points on the development of formal structures for fostering sound evaluation of education concern certification and licensing. Twenty years ago, there was no sign of such steps to control and assure the quality of evaluation work. Though the situation has not changed very much, there has been some movement. Notably, the state of Louisiana has instituted a program for certifying educational evaluators. At this point it is difficult to know if the Louisiana experience is only the tip of the iceberg, the first of many similar state programs for certifying educational evaluators.

In the foregoing analysis I think that I have put forward a fairly strong case that evaluation, though perhaps not fully mature, shows strong signs of an emerging profession.

FINAL COMMENTS ABOUT THE FUTURE AND AN AGENDA

I have no doubt that educational evaluation will continue. As we have been hearing recently (_____, _____, _____), education is about excellence, and it is a pervasive national concern. It needs to reach all the people, and it needs to be done well, much better than it is presently being done.

There will always be efforts to assure the quality of and to improve education. By definition, such efforts cannot go forward with any degree of effectiveness unless they are guided by sound evaluation.

Moreover, evaluation is now finding it necessary to interrelate its work in program evaluation and personnel evaluation. With the "crisis about excellence in the schools" at hand there is great pressure to evaluate educators' performance, and I believe evaluation has both an opportunity and a significant responsibility to bring about sound progress in this area.

A complex set of knowledges and skills are required to evaluate education effectively. Usually no one person possesses all the qualifications to do evaluations; hence they must be done by teams. Since there will never be enough specialists to do all the evaluations that are needed in education, there must be significant services to provide training and technical assistance to generalists in education. There also should be extensive efforts to build systematic evaluation procedures into curriculum materials. Of course there will continue to be much work to advance the theory and methodology of evaluation, and to promote the professionalization of the field. Clearly, there is a need for a sizeable and highly specialized work force in the evaluation field.

I believe that the federal government has an important responsibility to help advance evaluation's professionalization, and to assure that the field continues to improve. The government also has a responsibility to assure that members of minority groups are involved meaningfully in evaluating education, because public education is their main avenue to an improved station in life. Thus they have a very high stake in assuring

that educational opportunities for minority members are carefully and validly evaluated.

Basically, I have five recommendations for the U.S. Department of Education:

- Provide leadership and funds for training educational evaluators at the masters and doctoral levels. Training has waned. Many evaluators have left the education field. With the upturn in the economy and the responses to the crisis about excellence in the schools that are sure to follow, there is bound to be expanded development and a corresponding need for qualified evaluators.

- Provide special assistance for recruiting and supporting the training of minority persons in evaluation.

- Support an in-depth case study of the Louisiana experience in certifying evaluators. It is a unique and important case whose analysis should be highly instructive.

- Support research and development aimed at improving evaluators' abilities to identify and involve multiple clients and address their questions.

- Support systematic research and development work in the area of evaluation of educational personnel. There are many current projects to increase and improve evaluations of educational personnel; these should be closely studied to help assure that they will promote rather than thwart better teaching, administering, and learning.

I think the merits of these five recommendations should be obvious, given the foregoing analysis. Surely, the list is incomplete. But the recommendations are concrete. If implemented, the federal role in fostering sound evaluation of education would be strengthened greatly.

REFERENCES

AN ANALYSIS OF FIVE VIEWS OF EVALUATION

Eleanor Chelimsky

General Accounting Office

In this presentation I summarize and discuss the five preceding papers. I look at the ways each paper addressed the questions posed by Susan Klein, examine the substance of the answers given, and point out what the range of options seem to be. I close by offering a few comments of my own.

Basically, Susan posed three questions to the presenters: Where have we been in educational evaluation? Where are we now? And where are we going? She asked that answers to these questions be couched in terms of the five indicators described in her own paper.

On the basis of the answers supplied to the "where are we going" question the presenters were to decide whether educational evaluation is alive, dead, or undergoing reincarnation. As they did this, three of the presenters addressed all the questions and all the major indicators; another presenter examined past, present, and future with an emphasis on evaluators' status, visibility, and training; another presenter concentrated less on historical factors and more on the present and the future. In analyzing the various perspectives, then, it should be noted that the questions were not addressed to the same extent by each presenter.

How Did the Panel as a Whole Come Down on the Three Questions Posed?

Let's examine the three questions one by one and look both at individual views and at consensus or disagreement across panel members.

First: Where have we been in educational evaluation historically with regard to the five indicators? There seems to be pretty fair agreement

among those who addressed historical trends that we've had a pretty good past. However, the range of opinion moves from "scintillating" and "a truly remarkable chapter in the history of social science" (John Evans), through Sue Klein's moderate claim for "some evidence that evaluation is one effective educational improvement strategy," to Eva Baker's characterization of her experience as "both exhilarating and painful," with some reservations as to evaluation's usefulness unless it undergoes some reincarnation.

In support of their generally favorable views, presenters pointed to increases in financial support of evaluation (although Sue Klein and Eva Baker mentioned some discomfort with the lack of good data in this area), to increases in the number of evaluation studies, of evaluation units in government agencies at all levels, and of legislative mandates for evaluation, as well as to what Dan Stufflebeam calls "great expansion" in the professionalization of evaluation over the last 20 years.

In the opinion of most of the presenters, then, we've come a long way in the past two decades.

Second: Where are we today in educational evaluation? There seems to be a general feeling among the presenters that we're less well off than we used to be, but divergence of opinion as to why and to what degree.

For example, with regard to financial support, there were again the kinds of data problems (such as changed definitions over time and little information) prompting a range of views from "agencies with money seem to be doing more evaluation now than ever before" (Eva Baker), through Sue Klein's "investments have remained fairly stable" in some areas though others have been cut, to John Evans' "what was once a bountiful flow of

funds for the support of evaluation studies and staffs has dried up to a mere trickle."

On other indicators, Eva Baker says "the job market ... is fine," whereas John Evans speaks of " the virtual disappearance of job openings." Again, Eva Baker feels evaluation has continued to be institutionalized, although she points to the loss of "celebrity status" for the field.

Carl Sewell, however, doesn't believe much in the institutionalization of evaluation. He thinks such institutionalization is either lip service or politically manipulative. At the local level, he says, "the overriding concern" for doing evaluation is either "for satisfying requirements for continued funding," or else "to show how much ... schools have 'improved' since" a particular set of policymakers or administrators took charge. John Evans thinks educational evaluation is inextricably connected to social reform, is "not ... self-sustaining" in the face of counterreform and cutbacks, and thus has not been institutionalized, although some of its thinking may have been. Dan Stufflebeam believes that compared to evaluator training of the 60's and 70's, there has recently been a general decrease in efforts to recruit and train educational evaluators, especially minority evaluators.

While the overall feeling seems to be that, at best, educational evaluation is undergoing some problems, there are major differences of opinion about the causes of those problems. For example, Sue Klein and Eva Baker feel that evaluation has not always fulfilled its promise of usefulness and that its methodologies have not promoted evaluations of great use in improving education. From that position, both deduce a need for, and go on to advocate, federal support for more utility-focused research.

Carl Sewell sees a major problem for educational evaluation in administrators' failure to make "proactive use" of evaluation; that is, for policy formulation, for program planning, development, and implementation, and not just for accountability. From this position, he deduces the need to improve local administrators' skills in using evaluation to enhance the quality of their leadership and decisionmaking.

Dan Stufflebeam also sees problems in the utilization of evaluation, and adds to this his concern about the training of evaluators.

John Evans, on the other hand, believes that evaluation's problems are not related to any of the above. He believes that educational evaluation's decline is not due to any of the "numerous internal shortcomings and flaws ... which evaluators themselves have been so candidly self-critical about;" "nor ... to the fact that Congress and other levels of government were too politically motivated to utilize evaluation findings." Instead he feels that evaluation's problems are the result of counterreform. His thesis is that since it was the movement for social reform that brought "the extensive evaluation enterprise into being," it is the current cutbacks in reform programs that have brought about its decline.

Third: Whither educational evaluation? life, death, or reincarnation? One obvious but intriguing finding is that no presenter seems willing to predict death (although Eva Baker claims to support euthanasia for some forms of evaluation). Similarly, John Evans shrinks before the prospect and looks for "silver threads in the dark clouds." His advice to evaluators is that they should "weather these difficult times and rededicate themselves to the sustenance of this important enterprise."

Dan Stufflebeam has "no doubt that educational evaluation will continue." Both he and Eva Baker feel that the national concern about education -- expressed through commissions and press attention -- means that "the field still has some kick left in it" (Eva Baker). Carl Sewell thinks "evaluation ... will continue to be a prevalent practice at the local school district level," and Sue Klein also predicts "that evaluation will continue to live."

So, despite the different trends they identified, the different causes to which they attributed those trends, and the different needs they deduced from the problems they perceived, each presenter chose life and reincarnation over death. I would agree with the presenters on this issue, but like them, I have my own views of what happened, why it happened, and how to fix it.

Some Comments

First, I trace the origins of evaluation not only to social reform, but also to efforts during the 50's -- a period not especially celebrated for social reform -- to rationalize the management and resource allocation of defense missions and programs. Further, under the Nixon Administration -- again, not well known for its support of social reform but, on the other hand, concerned about and cognizant of management techniques -- evaluation flourished. If my analysis is correct, then, we may not have to wait for another reform cycle to get some renewed interest in evaluation, but only until an administration comes along that values good public management and sets out to do something about it, as did the Nixon Administration's Office of Management and Budget under Roy Ash.

Second, I would like to caution against too great a certainty about gloom-and-doom with regard to evaluation today, and underscore what Eva Baker and Sue Klein said about data uncertainties. The point is that some areas may be shrinking while others may be expanding. The Defense Department and GAO, for example, are expanding their evaluation or "program results" work. The Congress is asking for more and more evaluation and the departments will sooner or later be forced to respond. If the Department of Education survives efforts to abolish it and if education remains a major national issue, it seems to me that it's reasonable to expect a resurgence in educational evaluation. But even with more and better data on the U.S. evaluation effort, I think it will always be difficult to get real indicators of the magnitude and direction of the total investment in evaluation because of the pluralistic and generally ad hoc nature of our evaluation system.

I would note here that some of the most striking evaluation developments are currently occurring not in the United States but in Canada, where all governmental agencies are now required to perform evaluations cyclically of all major public programs. And these Canadian developments are not happening in an environment of social reform but rather in one of budget deficits (like ours), combined, however, with a widely shared governmental concern for improved public management.

Third, Carl Sewell's remark that local administrators seldom know how to use evaluation needs to be related to two other points: Eva Baker's observation that local control of educational services has resulted in evaluation demands that are too grandiose in their expectations relative to their funding and time constraints, and are also highly overspecified; and

to Dan Stufflebeam's discussion of problems in getting the Congress to specify its evaluation questions and his observation about the necessity to differentiate among audiences and to identify and directly address their different questions. If we connect these three points, they signify that problems in user knowledge of evaluation may exist at all levels of use and that, where they exist, they will directly affect the contributions that evaluation can make.

I think, then, that we need to put more of our energies into improving the linkages between evaluation and its users. We should be thinking more about different types of evaluation questions by different evaluation users at different points in time. And we need to address policy formulation questions as well as implementation and accountability questions.

My last point is that I am incurably optimistic about evaluation's future. I base this opinion on several factors. First, my experience with the Congress since coming to GAO has shown me (a) that it is perfectly possible to negotiate evaluative questions that are operationally defined and researchable and relevant to a user's information need; (b) that there is a growing market for neutral, objective information in the Congress; and (c) that members of Congress will take the trouble to look very carefully at data and methodologies involving issues about which they are concerned.

Second, the Canadian experience is likely to affect the United States' evaluation capacity sooner or later both by hiring away some of our evaluation professionals and by producing studies that will be considered in the oversight of our own programs.

Third, the Congress' increasing appetite for evaluation will surely spur more evaluation in the executive branch, if only because of the obvi-

ous countervailing power considerations. Finally, I see an analogy between evaluation's situation and that of budgetmaking. It took 53 years between the Budget and Accounting Act of 1921 setting up the Bureau of the Budget and the emergence of the Congressional Budget Office.

For all of these reasons, I believe that evaluation is not only very much alive now but that it is also destined for a bright future. It seems to me we can see evaluation's situation in one of two ways. We can look at the brief history of evaluation in public management, examine the short-term trends and whatever data are available on its present vitality, and decide it was a flash in the pan. Or we can take note of an important development in the direction of better public management, observe its increasing quality and use, and decide that while the field is still troubled with growing pains, it is highly likely to be successful in the end. My view is the latter one, of course, and I expect to see the long-term investment in evaluation result in the emergence of evaluation as a primary tool -- like budgetmaking, like auditing -- for the formulation, execution, and assessment of public policies and programs.

EDUCATIONAL EVALUATION WILL BE REINCARNATED AT THE STATE AND LOCAL LEVELS

Shirley A. Jackson

U.S. Department of Education

Eleanor Chelimsky's analysis of the presentations shows that most predict evaluation will live or be reincarnated. The audience for these presentations was also asked to make their predictions for evaluation by responding to a measure developed by Susan Klein and John Evans.

The measure requested participants to indicate their perception of past trends (1981-1984) and future prospects (1985-1988) in educational evaluation at the federal, state, and local levels in relation to the five indicators described in Sue Klein's paper. Participants judged whether each of the indicators had Decreased (D), Remained About the Same (R), or Increased (I). In addition, participants were asked to indicate whether they felt that educational evaluation is generally healthy, dying, or undergoing reincarnation. Responses indicate that a majority predict that educational evaluation is destined for reincarnation at the state and local levels.

Audience Response

Although the audience exceeded the room capacity (200), only about 30 people filled out and returned the forms. Moreover, some did not complete all of the items and so the number of responses reported in a cell does not always equal the total possible.

Figure 1 summarizes the participants' responses.

Figure 1

Predictions on the Future of Educational Evaluation

INDICATORS		TRENDS 1981-84 Education Levels			TRENDS 1985-88 Education Levels		
		Federal/ National	State	Local	Federal/ National	State	Local
1. Financial Support of Evaluation	D*	D=20	D=12	D=11	D=9	D=8	D=5
	R	R=3	R=4	R=6	R=11	R=10	R=12
	I	I=0	I=10	I=9	I=4	I=8	I=9
2. Official Status, Mandate for Evaluation, or Institutionalization	D	D=13	D=8	D=7	D=8	D=3	D=1
	R	R=5	R=4	R=8	R=12	R=15	R=14
	I	I=3	I=10	I=5	I=2	I=9	I=11
3. Status and Visibility of Evaluators	D	D=13	D=9	D=5	D=5	D=5	D=3
	R	R=8	R=9	R=11	R=14	R=13	R=15
	I	I=2	I=5	I=5	I=5	I=9	I=10
4. Prevalence, Visibility, and Extent of Evaluation of Activities	D	D=13	D=5	D=5	D=4	D=3	D=3
	R	R=3	R=6	R=4	R=13	R=14	R=9
	I	I=6	I=11	I=10	I=6	I=11	I=15
5. Value, Effectiveness, or Utility of Evaluation in Improving Education via decisions, knowledge, or problem solving skills	D	D=9	D=3	D=4	D=3	D=6	D=3
	R	R=8	R=13	R=9	R=15	R=13	R=10
	I	I=5	I=5	I=7	I=3	I=8	I=11

GENERAL STATE OF EDUCATIONAL EVALUATION:

Healthy = 0 Dying = 2 Undergoing Reincarnation = 11

*D - Decreased
R - Remained About the Same
I - Increased

Federal/National Trends

Most respondents agree that federal trends during 1981 to 1984 are typified by decreases on all five indicators, with a slightly healthier picture suggested for evaluation's value or effectiveness. Similarly, most respondents believe that current federal reductions on each indicator will remain in effect or be even more severe in the future.

State Trends

Respondents noted a different pattern of trends at the state level during 1981 to 1984. While about half see state-level reductions in financial support, almost as many respondents observe an increase. About half the respondents likewise note an increase in the status and prevalence of evaluation (the remaining half are roughly split between those noting reductions and those observing relative constancy). In terms of evaluators' status and visibility, respondents are roughly split among those noting state-level decreases, those opting for constancy, and those observing increases. A majority feel the value and effectiveness of evaluation has remained about the same during 1981 to 1984.

With respect to projected future trends on the five indicators, most respondents believe that current levels of state support will remain the same or be augmented even further.

Local Trends

While a significant proportion of respondents note decreases in financial support for evaluation at the local level, a majority observe increases in its prevalence and visibility from 1981 to 1984. Respondents were also relatively positive about progress on the other three indicators. They were most optimistic about future trends at the local level.

On the basis of these findings, it is hardly surprising that most of the respondents believe that educational evaluation is destined to undergo reincarnation.

Implications of Findings

These results indicate that evaluation at the federal level will be decreasing, while efforts at the state and local levels are expected to increase. If these trends are accurate, it seems clear that the future agenda for educational evaluation will shift from the federal/national to the state and local levels.

If this shift takes place, and certainly it does appear that federal funds will most likely continue to shrink, then continued funding and public support of theoretical and applied research will be heavily dependent upon the ability of researchers and evaluators to persuade decision makers at the local and state levels that their work can yield practical results for educational improvement. In order to continue and/or expand large-scale, sustained, and worthwhile research and evaluation efforts, researchers and evaluators will probably have to forge creative alliances with others in the public and private sectors.

The federal funding trend focuses educational resources at the state and local educational levels, which severely reduces discretionary/categorical dispersal of funds to universities. It will therefore be important for researchers and evaluators to involve state departments of education and local school systems in collaborative research and evaluation efforts. However, forging collaborative efforts will require changes in attitudes, increased trust, and improved working relationships between researchers/evaluators and practitioners.

Conclusion

In order for a reincarnation of educational evaluation to occur at the state and local levels, educational evaluators must become skillful entrepreneurs and adept at developing collaborative working relationships with state and local educational policymakers, program implementers, and evaluation coordinators/directors.

In addition to collaboration, evaluators must make their work user-friendly and direct their findings toward local needs and information uses. It seems to me, and it very likely would also appear so to our respondents, that such a collaborative, user-friendly, and user-driven reincarnation would be a healthy form for evaluation.