

19975

PROGRAM EVALUATION FOR THE '80's: DOING WITH LESS EFFECTIVELY

(Panel on Evaluation and Cost-Benefit Analysis at the Symposium of Washington Operations Research/Management Science Council, Washington D.C., November 2, 1981)

by Keith E. Marvin, Special Assistant to the Assistant Comptroller General for Program Evaluation, U.S. General Accounting Office

It is clear that major changes are coming in the Federal government's involvement in many facets of our economy and our society. Fast growth in Federal programs will be exceptions. Most programs will be doing well to hold their own and many will be cut or dropped. Program/evaluation became a function in nearly all Federal departments and many of the agencies during earlier growth phases of many of these programs. Since program evaluation is a support service to management and oversight officials, it will need to adjust to these new conditions to survive.

Looking at the future of programs or their evaluation function is for most of us like driving in a heavy fog. We feel there is a road out there but we can't see the curves, hills and valleys very clearly. In this fuzzy context, I do have some comments on some of the sign posts I expect to see coming out of the fog.

Although no one seems to have a precise accounting for evaluation expenditures by the Federal government, and even less for state expenditures in total, we have some feel for it. Various surveys by the Office of Management and Budget(OMB) and GAO showed rapid growth in Federal expenditures in the early 70's to about \$250 million annually by the mid 70's for non-defense departments and agencies. The most recent survey

019147

by GAO of FY 80 expenditures indicated that moderate decline had already begun, and I have heard nothing to suggest that this has not continued in FY 81. For example, one evaluation function has dropped its average cost per evaluation from \$375 thousand in 1975 to \$145 thousand, in 1980. Cuts like that are generally not possible by increased management efficiency alone. They require such things as reducing the number of local sites covered in the data collected, and by interviewing smaller numbers of clients at each site. Such changes weaken any generalizations and require the evaluator to take greater risks in drawing any conclusions. However, studies of use of evaluation in policy have not shown that methodology was a major factor compared to responsiveness and timeliness. Weakening of designs may not reduce use or demand significantly. Just "making it" or at least meeting demand for evaluation in an era of declining resources is the subject of this talk.

I see two components in a successful future for an evaluation function. The first, and probably the most essential for survival of the function, is helping managers to know where to cut programs effectively. The second, the "frosting on the cake" is improving evaluation management.

I believe the most successful evaluators will be those who show managers how to use better techniques to make the necessary cuts with the least loss of services from the program. This requires that managers know how to preserve the most efficient activities. One such technique may be Data Envelopment Analysis, a procedure based on linear programming principles and which provides relative measures of efficiency for multi-input, multi-output organizations. The technique works as follows: Some raters, either the evaluators or the managers or both, attach weights to the various outputs. Based on the prioritization yielded by these weights, DEA then classifies all activities in the program into one of two groups, efficient and inefficient. All efficient units receive a score of 1.0 while those judged inefficient are scored from 0 to 1.0 indicating how inefficient they are. The technique allows the evaluator to make a number of passes through the data, each with different priorities on the outputs. This can indicate the trade-offs in efficiency realized if priorities are shifted. Additionally, the technique produces a statistic which indicates, for those units in the efficient category, the magnitude of decrease in outputs necessary before they drop to the inefficient class.

Both the relative efficiency ratings and the size of the efficiency "buffer" should be useful to managers forced to allocate reduced resources. Managers can cut the inefficient units, cut those whose efficiency would remain high even after a cut, or adopt any combination of these cuts.

Evaluation management has several dimensions, all of which should be considered for improvement. First, the model of the function in each agency should be reviewed and improved as needed so its objectives are in tune with the people it serves. Second, staff management and use of contractors and consultants can be improved. With regard to consultant services there is some impatience in the Congress and consultant reform and disclosure bills are being considered in both the House of Representatives and in the Senate. GAO normally does not support legislative action to remedy problems that should be resolved administratively. However, since executive branch agencies, with few exceptions, have not acted administratively, GAO believes congressional action is necessary. A personal view on this is that sharing more of experience as to what works best can be an important aid. The inter-agency evaluation officials' group could take up this challenge.

The third area of evaluation management which can be improved is technical. It requires that new techniques and computational methods be exploited intelligently and in very practical ways. For example, if the budget won't support a reliable statistical design, then it may be best to use a case study approach. We have some effort going at GAO to see what improved case study methods have been tested recently. A very important area to exploit will be methods for second order analysis, i.e., using as the data base various data that others have already paid for. Several of these methods have been

5

around for ten years or so and have been tested well enough to provide some guidance. They include things called "found data", meta-analysis, synthesis of evaluations, re-analysis, and content analysis. The volume of material available is indicated by the three directories published by GAO covering Federal evaluations completed in 7 fiscal years, 1973-79, containing 5610 entries. Through reanalysis and reorganizing of the extensive data archives behind some of these studies new data has in effect been found. A variety of approaches for combining the results of prior evaluations has been characterized as sythesis, but normally it would be based on some organizing principles in terms of the issues of interest, the effectiveness measures common to the studies, etc. More specific techniques of reanalysis, meta-analysis, and content analysis can be used in appropriate situations, as follows:

- Reanalysis usually focuses on a single salient study or small number of studies. The original data is obtained and reworked with what are now believed to be better statistical methods or with different hypotheses.
- Meta-analysis is useful for combining the findings of a large number of studies statistically where a wide variation in/<sup>a particular</sup>effect has been found by these studies.  
Normally this would/<sup>not</sup>include going back to original data.
- Content analysis offers the potential of systematically analyzing written material in contrast to the emphasis of other methods on analyzing quantitative data.

Since more than half of Federal evaluation dollars are spent externally, better management must involve improving this area. For example, RFPs which specify only total dollars and time ignore the substantive issue under investigation. The baseline period in which data should be collected before the program change is implemented vs. the length of the implementation period; the selection of sites to be included; and such things as sample size in relation to population at the site might all have an important bearing on the efficiency with which the design obtains useful information. Techniques such as Bayesian statistics should be exploited to improve the confidence in design decisions.

To be most efficient evaluation designs may need to be changed as the data collection proceeds. For example, tests of statistical significance can indicate which of the questions need more data. If answers to a question indicate strong and uniform satisfaction on a point after the first few sites, the tests can indicate that not much more is likely to be learned by spending more data collection effort on that point. This might suggest stopping or reallocating the effort to more probing questions on points which do not indicate uniform satisfaction. Such things are more difficult procedurally under procurement practices which have never been tailored to procuring evaluation but perhaps budget constraints will be the incentive lacking before.

Most of the experience on which I based this statement is at the Federal level. However, it is fairly clear from activities of leaders in the evaluation section of the National Conference of State Legislatures and from a number of direct contacts GAO has had with state people that capability has grown rapidly. They will be faced also with severe budget constraints but the skills they have developed will I believe enable them to make some of the most exciting breakthroughs in the next ten years. They also will be ideally situated to make effective use of the mini- and micro-computer technology which is now practical for small offices and even home use. The small really creative analyses have been badly constrained by the limits and costs of large computers, and availability as well as costs of time shared systems. The revolution taking place in the computer business opens up fantastic opportunities for creative work, particularly at state and local levels.

In summary, I believe the budget constraints on both programs we evaluate and on our own evaluation resources can be the cause for advancing evaluation, not killing it. Demand for good timely evaluation will increase and opportunities for sound and creative evaluation management should be excellent.

## BIBLIOGRAPHY

- Charnes, A. and W. W. Cooper. Goal Programming and Multiple Objective Optimizations. Graduate School of Business Administration, Harvard University, Cambridge, Reprint Series (1977).
- Charnes, A., W. W. Cooper, and E. Rhodes. Measuring the Efficiency of Decisionmaking Units. European Journal of Operational Research, North Holland Publishing Co., Amsterdam (1978) 429-444.
- Lewin, Arie Y. and Richard C. Morey. Studies to Validate and Extend the Data Envelopment Analysis Method for Measuring Relative Efficiency of Decisionmaking Units. Final Report, Office of Naval Research, Washington, D. C. Grant No. N00014-80-C-0582. October, 1980.
- Boruch, Robert F. ed. Secondary Analysis. Jossey-Bass, San Francisco, CA. 1978.
- Smith, M. L. and Glass, G. V. Meta-analysis of Psycho-therapy Outcome Studies. American Psychologist, 1977, 32, 752-760.
- Krippendorff, K. Content Analysis: An Introduction to Its Methodology. Sage Publications, Beverly Hills, CA. 1980.